1996

A time series analysis of U.S. Army enlisted force loss rates

DeWald, Edward T.

Monterey, California. Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

THESIS

A TIME SERIES ANALYSIS OF U.S. ARMY ENLISTED FORCE LOSS RATES

by

Edward T. DeWald

September, 1996

Thesis Advisor: Robert R. Read

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE September 1996	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE A TIME SERIES ANALYSIS OF U.S. ARMY ENLISTED FORCE LOSS RATES

5. FUNDING NUMBERS

6. AUTHOR(S) Edward T. DeWald

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

The analysis and prediction of personnel loss behavior is critical to effective manpower planning and to the U.S. Army's Enlisted Personnel Strength Management System (EPSMS). In support of efforts to modernize the EPSMS, this thesis examines the method by which the Enlisted Loss Inventory Model (ELIM) analyzes loss rates and forecasts them into the future.

Time series analysis techniques seek to identify patterns in data and forecast them into the future via time based extrapolations. Four such methods were used to construct loss rate forecasts from data. These methods were the arithmetic mean, exponential smoothing (the current ELIM method), seasonal exponential smoothing and an autoregressive moving average model. Forecasted rates were used to project force strengths which were in fact known. The resulting errors in forecasted strength were analyzed, compared and contrasted with respect to the methods.

Error analysis revealed no significant performance differences between the methods. Hence, the simplest methods (mean and exponential smoothing) may be viewed as more economical and preferred.

14. SUBJECT TERMS Time Series Analysis, Forecasting, ELIM-COMPLIP, Loss Rates, Exponential Smoothing, Winter's Method, Box-Jenkins, ARIMA.

15. NUMBER OF PAGES 107

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18 298-102

Approved for public release; distribution is unlimited.

A TIME SERIES ANALYSIS OF U.S. ARMY ENLISTED FORCE LOSS RATES

Edward T. DeWald
Captain, United States Marine Corps
B.S., United States Naval Academy, 1990

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1996

ABSTRACT

The analysis and prediction of personnel loss behavior is critical to effective manpower planning and to the U.S. Army's Enlisted Personnel Strength Management System (EPSMS). In support of efforts to modernize the EPSMS, this thesis examines the method by which the Enlisted Loss Inventory Model (ELIM) analyzes loss rates and forecasts them into the future.

Time series analysis techniques seek to identify patterns in data and forecast them into the future via time based extrapolations. Four such methods were used to construct loss rate forecasts from data. These methods were the arithmetic mean, exponential smoothing (the current ELIM method), seasonal exponential smoothing and an autoregressive moving average model. Forecasted rates were used to project force strengths which were in fact known. The resulting errors in forecasted strength were analyzed, compared and contrasted with respect to the methods.

Error analysis revealed no significant performance differences between the methods. Hence, the simplest methods (mean and exponential smoothing) may be viewed as more economical and preferred.

TABLE OF CONTENTS

I. INTRODUCTION 1

A. BACKGROUND 1

B. THE SIGNIFICANCE OF LOSS RATES 2

C. THESIS OBJECTIVE AND ORGANIZATION 4

1. Objective 4

2. Organization 4

II. THE ELIM-COMPLIP SYSTEM 7

A. FUNCTIONALITY AND USE 7

B. INPUTS 8

C. STRUCTURE AND PROCESSING OVERVIEW 8

1. Characteristic Group Designator (CGD) 9

2. Rate Factor Generator (RFG) 11

3. Inventory Prediction Module (IPM) 11

4. Optimization Module (OM) 11

5. Report Generator Module (RGM) 12

D. ACCURACY 12

III. METHODOLOGY 13

A. TIME SERIES ANALYSIS 13

B. HISTORICAL TIME SERIES LOSS DATA 13

1. Partitioning the Data into Characteristic Groups 14

2. Service Lifetimes Calculated from Gain/Loss Records 14

3. Loss Rates from Lifetimes 15

4. The Time Series Data Template 15

C. TIME SERIES ANALYSIS AND FORECAST METHODS 17

1. The Mean Loss Rate 17

2. Simple Exponential Smoothing 18

3. Seasonal Exponential Smoothing 20

4. Auto-Regressive Moving Average (ARMA) 22

D. ANALYSIS OF FORECAST ERRORS 26

1. Measures of Effectiveness 27

a. Forecasted Strength 27

b. Monthly Relative Error (PE) in Forecasted Strength . . . 28

c. MIS Mean Relative Error (MRE) 29

d. Grand Mean Percent Error 30

2. Displays 30

a. Histograms 30

(1) Grand Histograms. 30

(2) Paired (Calender Year, YOS) Histograms. 31

b. Boxplots 32

(1) Grand Boxplots. 32

(2) Yearly Boxplots. 32

(3) Year of Service (YOS) Boxplots. 32

c. Line Graphs 32

IV. RESULTS 33

A. OVERVIEW 33

B. THE DISTRIBUTION OF ERRORS 33

1. Histograms 33

2. Boxplots 35

C. MEASURES OF EFFECTIVENESS 38

1. Estimated First Term Force Strength 38

2. Monthly Relative Errors in Estimated Strength 39

3. MIS Mean Relative Errors 40

D. POLICY EFFECTS ON THE RESULTS 41

E. SUMMARY 44

V. CONCLUSION 45

A. SUMMARY 45

B. OVERALL CONCLUSIONS 46

C. RECOMMENDATIONS FOR FURTHER STUDY 46

APPENDIX A. CGD & SLC PROGRAM (SAS) 49

APPENDIX B. TSDG PROGRAM (PASCAL) 53

APPENDIX C. EXPONENTIAL SMOOTHING FUNCTIONS 57

APPENDIX D. SEASONAL EXPONENTIAL SMOOTHING 59

APPENDIX E. ARMA FUNCTION 61

APPENDIX F. ERROR HISTOGRAMS 63

APPENDIX G. ERROR BOX PLOTS 73

APPENDIX H. TOTAL MONTHLY ACCESSION PLOTS 85

LIST OF REFERENCES 89

INITIAL DISTRIBUTION LIST 91

EXECUTIVE SUMMARY

The analysis and prediction of personnel loss behavior is critical to effective manpower planning and to the U.S. Army's Enlisted Personnel Strength Management System (EPSMS). In support of efforts to modernize the EPSMS, this thesis examines the method by which the Enlisted Loss Inventory Model (ELIM) analyzes loss rates and forecasts them into the future.

Monthly historical loss rates were constructed from personnel loss/gain event records. For each cohort under study, these rates represented the proportion of soldiers that left the Army during each month of service throughout their first term enlistment. The study included only those soldiers belonging to C-Group 1, and only while serving in there first term. The results of the study must always be caveated by this C-group 1 restriction, but they remain valid and important since over 45% of the Army's total accessions during the study period (1983 - 1994) were C-Group 1 soldiers.

All time series analysis techniques seek to identify patterns in the data and forecast them into the future via time based extrapolations. The methods examined in this thesis were the arithmetic mean, exponential smoothing (the current ELIM method), seasonal exponential smoothing, and an autoregressive moving average model. An analysis data set, containing loss rates from cohorts entering service between January 1983 and December 1988, was used to construct future rate forecasts.

Forecasted loss rates were used to construct monthly first term force strength projections six years beyond the last data month - from January 1989 to December 1994. The forecasts were then compared to known force strengths for the same periods. Comparisons were quantified and summarized by relative errors in forecasted strength. The errors were

displayed in a variety of forms to allow performance comparisons between the loss rate forecasting methods.

The analysis of errors in forecasted strength revealed no significant performance differences between the loss rate forecasting methods. The methods' error distributions were remarkably similar and all methods performed similarly with respect to world events and policies that affected first term force strength. In terms of complexity and sophistication, the methods rank from simplest to most complex according to; mean, exponential smoothing, seasonal exponential smoothing, autoregressive moving average. In terms of the mean percent error in forecasted strength, the methods rank in the order of best to worst according to; exponential smoothing (0.55%), mean (1.83%), autoregressive moving average (1.84%), seasonal exponential smoothing (2.40%). While these mean percent errors are useful and contribute to the overall evaluation, they obscure the unique behavior of each method and may not be used to definitively identify any one method as superior to another.

Since no significant performance differences may be noted between the methods, the simplest methods may be viewed as more economical, and thus favored. Accordingly, the current ELIM-COMPLIP method, exponential smoothing, has been validated with respect to the other methods, but selection of the smoothing constant remains an analytical dilemma without precise interpretation.

Another interesting result is the viability of the arithmetic mean as an estimate of loss. Simple, understandable and effective, the arithmetic mean of past loss rates proved itself as a valuable forecast that could facilitate timely answers to many manpower planning problems.

Capable of extension beyond the scope of this thesis, the autoregressive moving average method is worthy of further study. Able to analytically incorporate other variables which may affect loss behavior, the model may achieve greater accuracy than demonstrated here. Such a study will require several more years of loss data than was available for this thesis, an accurate record of strength affecting events and policies, and data sets containing econometric variables which may effect loss. Any future study of the autoregressive moving average method should also consider examination of lifetime regression and survival analysis techniques, as they too are capable of incorporating the effects of other variables into forecasts.

I. INTRODUCTION

A. BACKGROUND

Manpower is the power, in terms of people, available or required for work or military service (Webster, 1992). As such, there are two competing elements to any manpower problem - the number of people available and the number of people required. In most organizations these quantities are dynamic. They change with the size, demography, and inherent traits of the supply population, and also with the size, structure, and objectives of the target organization. As a result, organizations must actively engage in manpower planning and analysis to achieve an efficient pairing of the personnel supply and the required work force. Necessarily, this planning is focused on the unknown future, and often based on the statistical analysis of the past.

The U.S. Army is actively engaged in manpower planning and analysis. The Army IS people. Soldiers are recruited from the American population, trained, organized, and equipped for the service and defense of our nation. The link between the Army's operational competence and the effectiveness of its manpower planning and analysis is undeniable - and unmatched by other nonmilitary organizations. Recognizing this in the early 1970's, the Army developed an integrated series of computer based mathematical models to meet post-Vietnam conflict demands for improved manpower planning and budgeting. A main component of this system is the Enlisted Loss Inventory Model - Computations of Manpower Programs using Linear Programming (ELIM-COMPLIP). The model is the cornerstone of the Army's Enlisted Personnel Strength Management System (EPSMS).

The original ELIM-COMPLIP system has not kept pace with recent advances in computing technologies and analytical methods. Accordingly, in 1995, the Army's Office of the Deputy Chief of Staff for Personnel (ODCSPER) initiated a comprehensive redesign effort to modernize the EPSMS. In support of this effort, this thesis examines the method by which important ELIM-COMPLIP parameters are estimated. Specifically, these parameters are the *loss rates* which forecast the proportion of soldiers that will leave the Army during each month within a given planning horizon.

B. THE SIGNIFICANCE OF LOSS RATES

A simple manpower system may be viewed as a series of personnel flows between the personnel supply and the available employment positions. In Army terms, the personnel supply is the American population[1] and the available positions are determined by the Army's Force Structure Allowance (FSA) and Tables of Organization (T/O). Figure 1.1 contains an elementary schematic of a manpower system. Notice, flows from the population to the work force are called *Accessions*, while the flows in the opposite direction are called *Losses*.

Accessions and losses are not the only personnel flows in a complex manpower system. Additional flows are usually present and reflect promotions, demotions, and other changes in employment status. Despite the presence and influence of these flows, losses remain the most fundamental quantity for manpower planning, (Bartholomew and Forbes, 1979). Losses arise from individual decisions to leave the work force, retirements, dismissals,

---

[1]Omitted from this point is the obvious fact that the personnel supply is more narrowly defined as those individuals who qualify for military service. Suffice it to say, individuals must possess certain prerequisite traits and characteristics, and must meet established mental, physical, and behavioral standards.

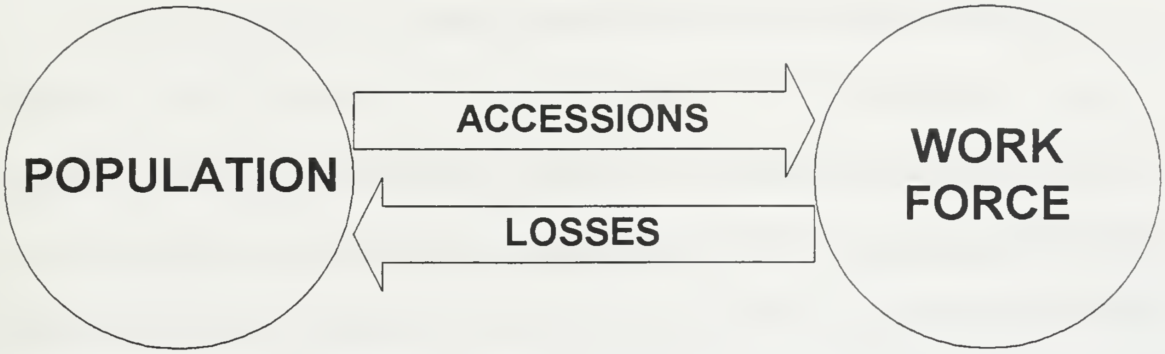

Figure 1.1 A simple manpower system.

disabilities and deaths. In general, losses cannot be controlled by leaders and managers. Furthermore, losses create vacancies in the work force and thus provide opportunities for others to advance and new recruits to join. Accordingly, successful manpower planning and analysis depends on the ability to describe and predict patterns of loss.

Traditionally, the analysis and prediction of loss behavior is accomplished via loss rates, (Bartholomew and Forbes, 1979). Empirically, a loss rate is simply the number of people who left service or employment, divided by the total number of people employed. While aggregate loss rates may be constructed in this way, separate loss rates may also be constructed for each loss type, and for each homogeneous subpopulation. Whether aggregate or separate, loss rates are constructed from historical data, statistically analyzed, and then forecast into the future. The accuracy of these forecasts then determines the reliability and validity of all subsequent manpower planning and analysis. Consequently, loss rates are critical parameters estimated in the U.S. Army's ELIM-COMPLIP system.

C. THESIS OBJECTIVE AND ORGANIZATION

1. Objective

The objective of this thesis was to conduct a historical time series analysis of U.S. Army enlisted manpower loss rates to identify the most accurate and appropriate time series forecasting methodology. En route to this objective, the following tasks were performed;

1. Monthly historical loss rates were constructed from a raw database containing individual accession and loss event records. The rates were calculated for homogeneous subpopulations of soldiers, without respect for cause of loss. The resulting time series data was then partitioned to produce analysis and validation data sets.

2. The current ELIM-COMPLIP forecast methodology, Exponential Smoothing (ES), was used to construct loss rate forecasts from the analysis data set. This was done to gain perspective on the current method's computational complexity and accuracy.

3. Other appropriate time series analysis methods were used to forecast loss rates from the analysis data set. These methods included the arithmetic mean, Seasonal Exponential Smoothing (SES), and an AutoRegressive Moving Average (ARMA) model.

4. Appropriate displays and measures were developed and used to evaluate each method's forecast error with respect to the validation data set.

5. The forecasting methodologies were compared and contrasted. Evaluations were made to identify the most accurate and appropriate method according to its computational complexity and demonstrated accuracy.

2. Organization

This introduction provides the reader with the motivation, objective, and organization of this thesis. Subsequent chapters will build on this foundation and provide the computational details and analytical results to satisfy the objective.

Chapter II contains an overview of the ELIM-COMPLIP system and is provided to enhance the reader's appreciation for the importance of loss rates in the EPSMS. Additionally, the chapter adds perspective and depth to many of the introductory points made.

Chapter III contains the computational details employed to accomplish the thesis objective. The chapter first describes the time series data template and how it was constructed from the source database. Next, the chapter describes each time series method used to forecast loss rates. Lastly, the quantities and displays used to analyze the forecast errors are defined and described.

Chapter IV presents and discusses the results obtained from each forecasting methodology. Finally, in Chapter V, I offer my conclusions and recommendations regarding the time series analysis and forecast of U.S. Army enlisted manpower loss rates.

II. THE ELIM-COMPLIP SYSTEM

A. FUNCTIONALITY AND USE

The Enlisted Loss Inventory Model (ELIM)[2] is an integrated system of computer-based mathematical models used by the US Army to model the strength of the enlisted force at the aggregate level.[3] Initially developed in the early 1970's by the Office of the Deputy Chief of Staff for Personnel (ODCSPER), the model is currently used and cosponsored by ODCSPER, the Office of the Assistant Secretary of the Army for Manpower and Reserve Affairs (OASA (M&RA)), and the Personnel Command (PERSCOM). It is directly managed and executed by the Director of Manpower, Military Strength Programs Division, ODCSPER. (Dillaber, 1996)

ELIM is designed to forecast the US Army's aggregated active force strength, gains and losses in the execution year, budget year, and the five years contained in the Five Year Defense Plan (FYDP). According to GRC (1989) the model performs the following listed tasks.

1. Forecasts enlisted losses, reenlistments, and required accessions based on historical behavior, or known, or contemplated policy.

2. Forecasts the total Active Army personnel strength inventory, consistent with the forecast of accessions, losses, and reenlistments.

---

[2]The model's actual designation is ELIM-COMPLIP (Enlisted Loss Inventory Model - Computations of Manpower Programs using Linear Programming). ELIM is widely accepted as a simpler means of reference to the model and will be used for the remainder of this thesis.

[3] ELIM is aggregated in that it does not model strength with respect to Military Occupational Specialty (MOS) or Ranks. MOSLS, the Military Occupational Specialty Level System, begins with the output from ELIM and models the disaggregate level.

4. Forecasts the enlisted Non-Prior-Service (NPS) accessions required to achieve the Active Army Force Structure Allowance.

5. Produces information required by US Army manpower managers and decision makers.

B. INPUTS

ELIM receives manual input, historical database updates and other model outputs. Manual inputs include the Force Structure Allowance (FSA) from the Deputy Chief of Staff for Operations (DCSOPS), the Notional Force Structure from the Deputy Chief of Staff for Personnel (DCSPER), user defined constraints and assumptions, and user selected analytical options. Historical database updates are provided monthly by the Military Personnel Information Systems Command (PERSINSCOM) and consist of extracts from the Active Army's Enlisted Master File (EMF) and the Gain/Loss Transactions File (GLF). Other models providing input to ELIM include the Trainees, Transients, Holdees and Students (TTHS) model and the Female Enlisted Loss Inventory Model (FELIM). TTHS forecasts the number of soldiers unavailable to the operating forces, while FELIM provides loss and inventory forecast for female soldiers. (Dillaber, 1996)

C. STRUCTURE AND PROCESSING OVERVIEW

ELIM contains many modules with specific functions. The modules most critical to understanding ELIM's structure and processing include the Characteristic Group Designator (CGD), Rate Factor Generator (RFG), Inventory Prediction Module (IPM), Optimization

Module (OM)[4], and the Report Generator Module (RGM). Figure 2.1 contains a simple diagram representing these modules and there interaction.

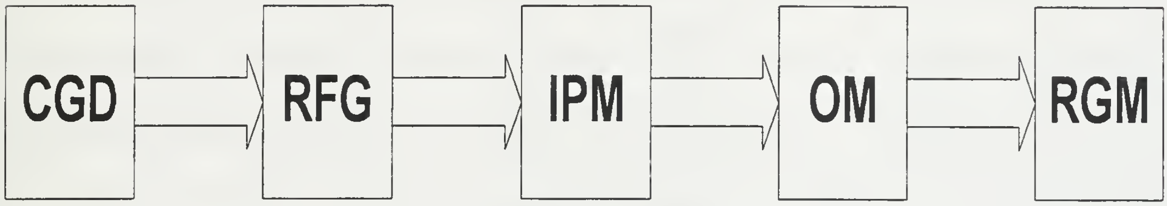

Figure 2.1 ELIM Modules

1. Characteristic Group Designator (CGD)

The CGD receives the EMF and GLF database updates and partitions the data into homogeneous subpopulations of soldiers in the active Army force. For soldiers in their first term of enlistment, homogeneity is assumed within Characteristic Groups (CG) determined by a soldier's gender, education, Armed Forces Qualification Test (AFQT) score, term of enlistment, and entry level training time. Table 2.1 summarizes the current Characteristic Group designations and traits. Soldiers beyond their first term of enlistment are identified as career level soldiers[5] and treated as a homogeneous subpopulation of their own. (GRC, 1989)

---

[4]The Optimization Module is really a combination of the Matrix Generator Module (MG) and the Linear Program Module (LP).

[5]Career Soldiers are those soldiers serving in their second or higher term of service, or who have successfully completed at least 55 months of service.

CG	Gender	Education	AFQT Category	Term	Training Time*
1	M	HSDG	I-IIIA	3,4	2-13
2	M	HSDG	IIIB	3,4	2-13
3	M	HSDG	IV-V	3,4	2-13
4	M	NHSDG	I-IIIA	3,4	2-13
5	M	NHSDG	IIIB-V	3,4	2-13
6	F	HSDG	I-IIIA	3,4	2-13
7	F	HSDG	IIIB-V	3,4	2-13
8	F	NHSDG	I-V	3,4	2-13
9	M	HSDG & NHSDG	I-V	2,5,6	2-13
10	F	HSDG & NHSDG	I-V	2,5,6	2-13

TABLE KEY:
CG: Characteristic Group
Gender: Male (M), Female (F)
Education: High School Degree (HSDG), No High School Degree (NHSDG)
AFQT Category: I-IIIA 99-50 percentile
IIIB 49-39 percentile
IV 30-21 percentile
V 0-20 percentile
Term: Length of enlistment contract, in years.
***Training Time:** months of entry level training received. Only tracked for soldiers on a variable enlistment length (VEL) contract that adds training time to the enlistment length. Program initiated in April 1985.

Table 2.1 Currently Defined Characteristic Groups (GRC, 1995)

2. Rate Factor Generator (RFG)

The RFG receives the partitioned data sets from the CGD and uses them in two component modules called the Qualitative RFG and the Non-Qualitative RFG. The Qualitative RFG forecasts first term loss rates based on historical loss activity and user defined analytic parameters. The Non-Qualitative RFG accepts career force data and user controls to forecast career level loss activity. (Dillaber, 1996)

3. Inventory Prediction Module (IPM)

The IPM uses the RFG's output, along with accession forecasts or goals, to calculate the projected force strength in any time period, according to the basic manpower accounting equation,

$$N_{t+1} = N_t - L_t + G_t \qquad \textbf{(2.1)}$$

where t indexes the time periods (typically months), N_t is the number of soldiers in the active Army at the beginning of period t (Force Strength), L_t is the number of soldiers that left service during period t (Losses), and G_t is the number of soldiers accessed during period t (Gains). (GRC, 1989)

4. Optimization Module (OM)

The OM receives IPM strength forecasts, TTHS and FELIM model outputs, and user supplied accession requirements, strength limitations, and force quality goals. Combining these data elements, constraints, and appropriate decisions variables for the analysis objective, the OM creates a large linear optimization data structure and seeks to minimize the Operating Strength Deviation (OSD) given by,

$$OSD_t = N_t - L_t + G_t - TTHS_t \quad \textbf{(2.2)}$$

where t, N_t, L_t, and G_t are as defined in equation (2.1), $TTHS_t$ is the forecasted number of Trainees, Transients, Holdees and Students in period t, and OSD_t is the Operating Strength Deviation for period t. (GRC, 1989)

5. Report Generator Module (RGM)

The RGM receives all useful output from the modules and produces a series of reports known as the Active Army Military Manpower Program (AAMMP). These reports are used to establish U.S. Army Recruiting Command (USAREC) operational recruiting missions, to determine the impact of manpower policies before and after they are implemented, and to plan the manpower budget. (Dillaber, 1989)

D. ACCURACY

GRC (1989) states "strength projections of ELIM have attained a level of accuracy within +/- 0.5% (for at least a 12-month horizon) of actual observation." Dillaber (1996) states, "the average error on loss projections is about 5% and the average error on man-year projections is about .1%." The precise meaning of these two statements is unclear, yet they appear to be the only testimonies to ELIM's accuracy. The author is unaware of any comprehensive examination of the model's forecast errors.

III. METHODOLOGY

A. TIME SERIES ANALYSIS

A times series is a list of observations paired with, and ordered by the time at which the observations were made. Time series analysis methods then seek to identify historical patterns in the data and forecast into the future via time-based extrapolations of those patterns, (Makridakis and Wheelwright, 1985). This thesis examines monthly personnel loss rates observed with respect to month in service (MIS). The remainder of this chapter presents a rigorous description and derivation of the time series data structure and analysis.

B. HISTORICAL TIME SERIES LOSS DATA

ELIM's working database is a merger of the monthly extracts from the Army's Enlisted Master File (EMF) and Gain/Loss Transaction File (GLF). The resulting file, called the Small Tracking File (STF), contains demographic information and gain/loss history on every non-prior service enlisted soldier accessed into the active Army during the last six years. The individual data records are further grouped with respect to each soldier's month of accession. Such a group, entering service at about the same time, is called a cohort, (Bartholomew and Forbes, 1979).

The database used for this thesis is the merger of two STF's. The resulting database contains one data record for every non-prior service soldier that entered the Army between January, 1983 (8301) and December, 1994 (9412). The database contains 1,066,413 records and required significant transformation to derive loss rate data. Detailed in the sections that follow, the transformation resulted in the Time Series Data Template (TSDT).

1. Partitioning the Data into Characteristic Groups

The Characteristic Group Designator and Service Life Calculator (CGD & SLC) listed and described in Appendix A, uses the STF's demographic information to partition the data into the C-Groups defined in Table 2.1. Partitioning revealed over 45 percent of the active Army accessions between 8301 and 9412 were C-Group 1 soldiers. Clearly, if a loss rate forecasting methodology were to be accepted as appropriate and accurate, it must be so with respect to C-Group 1 accessions. Seeking to capitalize on this idea, and in light of the number of C-Group partitions and research time constraints, only C-Group 1 loss rates were analyzed. All conclusions must be caveated by this fact.

2. Service Lifetimes Calculated from Gain/Loss Records

The CGD & SLC calculates each soldier's service lifetime from their Gain/Loss record. Recall from Chapter II, ELIM handles first term and career force loss rates separately, and only first term loss rates quantitatively. Accordingly, the lifetime of interest is each soldier's first term service lifetime. More specifically, a soldier's first term service lifetime is defined as the number of months in service from accession to the end of their first term, whether that ending was due to some type of loss, re-enlistment, or enlistment extension. Using this definition, it should be clear that if a soldier re-enlisted (or was discharged, or extended) during his or her 34th month in service, then that soldier's first term service life was 34 months.

A soldier's first term service lifetime was considered censored if there was not a term-ending event recorded in their Gain/Loss record prior to the last update of their gain/loss record, or prior to the maximum possible number of months in service for their enlistment

contract (48 months for C-Group 1 soldiers). The later censoring case was rare and most likely caused by data errors.

3. Loss Rates from Lifetimes

The Time Series Generator (TSG) listed in Appendix B processes the service lifetime data created by the CGD & SLC. The TSG calculates Kaplan-Meier estimates (Kalbfleisch and Prentice, 1980) of loss for every cohort *r*, in each month of service *s*. The loss rates are given by,

$$\lambda(r,s)=\frac{d(r,s)}{N(r,s)} \qquad \textbf{(3.1)}$$

where λ(r,s) is the loss rate for cohort *r* during month in service *s*, *d(r,s)* is the number of soldiers lost from cohort *r* during month in service *s*, and *N(r,s)* is the total number of soldiers from cohort *r* still in service at the beginning of their s^{th} month of service. A censored lifetime during month in service *s* contributes to $N(r,s)$, but not $d(r,s)$.

4. The Time Series Data Template

The Time Series Data Template (TSDT) is a transformation of the loss rates defined by cohort *(r)* and month in service *(s)*, to rates defined in terms of real time *(t)* and month in service *(s)* . The relation,

$$t = r + s - 1 \qquad \textbf{(3.2)}$$

holds, and is illustrated by soldiers belonging to cohort 8301 (*r*=1), in their fourth month in service (*s*=4), are serving in real time 8304 (t=4, when t=1 corresponds to 8301).

			Month of Service				
	Time Index	Time YYMM	$s = 1$	$s = 2$	$s = 3$	...	$s = S$
Analysis Data	$t = 1$	8301	$\lambda(1,1)$				
	$t = 2$	8302	$\lambda(2,1)$	$\lambda(2,2)$			
	$t = 3$	8303	$\lambda(3,1)$	$\lambda(3,2)$	$\lambda(3,3)$		
	⋮	⋮	⋮	⋮	⋮	⋱	
	$t = S$	8701	$\lambda(t,1)$	$\lambda(t,2)$	$\lambda(t,3)$	...	$\lambda(t{=}S,S)$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	$t = T'$	8812	$\lambda(T',1)$	$\lambda(T',2)$	$\lambda(T',3)$	...	$\lambda(T',1)$
Validation Data	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	$t = T$	9412	$\lambda(T,1)$	$\lambda(T,2)$	$\lambda(T,3)$	...	$\lambda(T,S)$

Figure 3.1 The Time Series Data Template (TSDT)

The TSDT displayed in Figure 3.2 is a (T x S) matrix where T equals the total number of real time months of data, and S is the maximum number of months in service for the C-Group's greatest term of enlistment. For C-Group 1 and the available data, T= 144 and S=48. The template is further divided by T' which is the index of the last month in the *analysis data set* used to derive estimates of loss. The *validation data set,* used to evaluate forecast errors, contains all months t, such that $t > T'$. For this thesis, $T' = 72$ which corresponds to 8812.

For the analysis of forecast errors it will be useful to index the TSDT by Calendar Year and Year of Service (YOS). Calendar Year i corresponds to the set of twelve rows index by t = $\{12(i - 1) + 1, 12(i - 1) + 2, ..., 12(i - 1) + 12\}$, where $i = \{1, 2, 3, ..., 12\}$ corresponds to years {1983, 1984, 1985, ..., 1994} respectively. Likewise, YOS j corresponds to the set of twelve columns indexed by $s = \{12(j - 1) + 1, 12(j - 1) + 2, ..., 12(j - 1) + 12\}$, where $j = \{1, 2, 3, 4\}$ corresponds to each of the four possible years of service.

A detailed examination of the TSDT and equation (3.2) reveals cohorts *(r)* remain constant along the diagonals of the template. Accordingly, there are no data elements above the $t=s$ diagonal containing the first observed cohort's loss rates. Lastly, notice each column of the template is a monthly time series of loss rates with respect to a fixed month in service.

C. TIME SERIES ANALYSIS AND FORECAST METHODS

1. The Mean Loss Rate

Naive forecasting methods are those approaches which provide forecasts without the use of sophisticated techniques. Generally, they are obtained with little effort, but still prove to be valuable forecasts. For this reason, naive forecasts serve well as the basis for comparing resultsfrom other forecasting methods. Obviously, any method must be worth its effort in terms of accuracy over the naive forecasts. This idea is often referred to as the rule of parsimony which may be more simply stated as, “keep things simple,” (Makridakis and Wheelwright, 1978).

The arithmetic mean of a time series is one such naive forecast of future behavior. Easily calculated and understood, the mean makes use of the available data, and it has a simplistic and familiar appeal. Accordingly, the mean of each time series was calculated and

used as the forecast for all future months within the time horizon. These forecasts, and their resulting errors, were then used as the basis for comparing the results of the remaining methods.

2. Simple Exponential Smoothing

Simple Exponential smoothing is the current method used by the ELIM system to forecast loss rates. The method averages past values of the time series (smoothing) in a decreasing (exponential) manner. Easy to understand and implement the method is suitable for forecasting time series data that fluctuate within a known range, with little to no growth or decay in value over time. The loss rate data was assumed to satisfy this form of stationarity.

To define the exponential smoothing method in detail, consider the time series $\{X_1, X_2, ..., X_t, ..., X_N\}$. Forecasts are obtained from the recursive relation,

$$F_{t+1}=F_t+\alpha(X_t-F_t) \tag{3.3}$$

where, beginning with t = 1, the forecast for the next time period (F_{t+1}) equals the forecast from the last time period (F_t), plus the weighted error of the last periods forecast. Forecasts are computed in this way for all t, up to and including t = N+1. The final value, F_{N+1} is then the exponentially smoothed forecast for all future periods. (Makridakis and Wheelwright, 1978)

The weighting term (α) is known as the exponential smoothing constant, dampening factor, or just simply alpha. Alpha values range between zero and one and function as a control for error. Weighting the importance of past errors in the current forecast, alpha places exponentially less weight on earlier errors according to the recursive relation of equation

(3.3). Obviously, alpha's value influences the accuracy of forecast and reflects some aspect of the time series' behavior. In general small alpha values smooth the data more than larger values, and they are particularly suited for data with considerable random fluctuations. In contrast, larger alpha values imply the best forecast is near the most recent observed value, and they are used for data with small random fluctuations or clear patterns. (Makridakis and Wheelwright, 1978)

In ELIM, alpha values for each MIS time series may be specified by the user or calculated to minimize the mean square error obtained by the application of equation (3.3). Formally, alpha is selected by solving the optimization problem,

$$\text{Minimize:} \quad \frac{1}{N-1}\sum_{t=2}^{t=N}(X_t - F_t)^2 \tag{3.4}$$

$$\text{Subject to:} \quad F_{t+1} = F_t + \alpha(X_t - F_t) \quad \forall\, \{t = 2, 3, \ldots, N\} \tag{3.5}$$

$$0 \le F_t \le 1 \qquad \forall\, t \tag{3.6}$$

$$0 \le X_t \le 1 \qquad \forall\, t \tag{3.7}$$

$$0 \le \alpha \le 1. \tag{3.8}$$

ELIM solves this problem numerically, incrementing alpha from 0 to 0.6 by steps of size 0.01, and choosing the alpha which achieves the smallest mean square error. GRC (1989) explains the reason alpha is only allowed to range from 0 to 0.6 as follows,

> "... [Alpha] has been restricted to lie between 0.0 and 0.6, since, if allowed to cover the full range, the optimization methodology tends to yield large values of alpha for many [of the month of service] time series, and in an unpredictable way. Clearly, such large values of alpha are counterintuitive so an arbitrary decision was made at the Army's request to restrict valid alphas to the 0.0 to 0.6 range."

Exponential smoothing forecasts were constructed using the current ELIM methodology and alpha range restrictions. Appendix C contains a description and listing of the computer programs written and used to obtain these forecasts.

3. Seasonal Exponential Smoothing

Seasonal exponential smoothing is similar in principle to simple exponential smoothing, but accounts for recurrent or seasonal patterns in the data. Recall from the previous discussion of simple exponential smoothing, large alphas were suitable for data with clear patterns in time. Also recall, the minimum mean square error alpha values tended to be large when left unrestricted. Together, these facts suggested the need to explore seasonal exponential smoothing to forecast loss rates.

Winters' two parameter seasonal exponential smoothing (Makridakis and Wheelwright, 1895) employs two smoothing equations, each with its own smoothing constant, and one forecast equation. The method also introduces seasonal indices which are similar to those found in many econometrics applications which adjust forecasts for seasonal effects. The three equations used in Winters' method are

$$S_t = \alpha \frac{X_t}{I_{t-L}} + (1-\alpha)S_{t-1} \tag{3.9}$$

$$I_t = \gamma \frac{X_t}{S_t} + (1 - \gamma)I_{t-L} \tag{3.10}$$

$$F_t = S_t \, I_{t-L} \tag{3.11}$$

where S is the smoothed value of the deseasonalized series, I is the smoothed value of the seasonalized factor or index, L is number of time periods in a complete cycle (e.g. L=12

months in a year), F is the forecast value, α is the deseasonalized series smoothing constant, and γ is the seasonal index smoothing constant. (Makridakis and Wheelwright 1985)

Winters' method is less intuitive than simple exponential smoothing and it requires a more complicated initialization procedure. To illustrate the technique and document the initialization policy, consider a series of N data values represented by X, indexed by t, and containing a seasonal pattern repeated every L periods. To allow for the I_{t-L} values required by equations (3.9) and (3.10), it is necessary to begin the time indices (t) one complete cycle prior to the first observation. Accordingly, the first data value becomes X_{L+1}, and the last X_{N+L}. The initialization policy then sets $S_{L+1} = X_{L+1}$ and $I_1 = I_2 = \ldots = I_{L+1} = 1$.[6] Next, equations (3.9) - (3.11) are evaluated for all t = {L+1, L+2, ..., N+2L}. The values F_{N+L+1}, F_{N+L+2}, ..., F_{N+2L} are the forecasts for the next entire cycle (L periods) into the future. Additionally, these L forecasts are used as the forecast for all cycles within the planning horizon.

As is the case with simple exponential smoothing, picking appropriate values for the smoothing constants alpha (α) and gamma (γ) is critical to obtaining accurate forecasts. Accordingly, a similar minimum-mean-square-error method was used to determine alpha and gamma for each month of service time series. Since there were two unknown parameters this endeavor proved to be much more computationally intensive and time consuming. To mitigate these effects, alpha and gamma were restricted between 0.0 and 0.6, and only determined to the 0.02 accuracy level. Analytical justification for the range restriction is

---

[6] For any complete cycle (L periods) the sum of the I_t values will equal L. Computationally, this is often disturbed by round off errors, thus making it necessary to re-normalize a cycle's indices after the complete cycle is calculated.

identical to the argument presented in defense of the simple exponential smoothing restrictions on alpha. The specified accuracy level was necessary to achieve reasonable computational time.

Appendix D contains a description and listing of the computer programs written and used to obtain Winters' seasonal exponential smoothing forecasts of loss rates.

4. Auto-Regressive Moving Average (ARMA)

All time series forecasting techniques are based on the belief that future observations may be expressed as a function of past values and patterns. The methods described thus far apply this principle with a fixed (once specified) weighting scheme applied to historic data and trends. The autoregressive moving average (ARMA) method does not use a fixed weighting scheme, but instead seeks optimal weights for data included in the model. The ARMA method has two components - an autoregressive (AR) and a moving average (MA) component. Further, the method requires the time series to be stationary.[7] (Makridakis and Wheelwright, 1978)

Equation (3.12) is called an autoregressive scheme. Expressing future values as a linear combination of past ones, the relation is similar to a regression equation with past observations as the independent variables, and the forecast as the dependent variable. The subscript p specifies the *degree* of the autoregression and determines the number of past values included in the relation. The coefficients (ϕ_1, ϕ_2, ..., ϕ_p) are the regression parameters

---

[7] While strict ARMA models require the time series to be stationary, closely related AutoRegresssive Integrated Moving Average (ARIMA) models accept non-stationary time series data and transform them to stationarity using a technique called differencing. Since the loss rate data was sufficiently stationary, these models were not explored. For further reading on the subject of ARIMA models see Box and Jenkins (1976)

or weights, and ε_t represents the unpredictable randomness of the process in period t. (Makridakis and Wheelwright, 1978)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \ldots + \phi_p X_{t-p} + \varepsilon_t \qquad \textbf{(3.12)}$$

Equation (3.13) is called a moving average scheme. The moving average component of an ARMA model expresses future values as a linear combination of past forecast errors. The subscript q represents the *degree* of the moving average process and determines the number of errors to include in the relation. The coefficients (θ_1, θ_2, ..., θ_q) are the parameters or weights of the moving average process, and the ε_{t-i}'s represent the difference between the forecast and observed value in period *t-i*. (Makridakis and Wheelwright, 1978)

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_{30} \varepsilon_{t-3} - \ldots - \theta_q \varepsilon_{t-q} \qquad \textbf{(3.13)}$$

A mixed autoregressive moving average model is the result of combining equations (3.12) and (3.13). The resulting ARMA equation of order (p, q) is given in equation (3.14). (Makridakis and Wheelwright, 1978)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_P X_{t-P} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \qquad \textbf{(3.14)}$$

When the data exhibit seasonality, both the AR and MA schemes of equations (3.12) and (3.13) may not be sufficient. Intuitively, one would expect a better forecast for the current period if it were constructed as a function of values from the same period of past cycles. For example, consider data with a seasonality of length L. Equation (3.15) is the appropriate AR scheme of degree P, and equation (3.16) the appropriate MA scheme of degree Q.

$$X_t = \phi_L X_{t-L} + \phi_{2L} X_{t-2L} + \phi_{3L} X_{t-3L} + \ldots + \phi_{PL} X_{t-PL} + \varepsilon_t \qquad \textbf{(3.15)}$$

$$X_t = \varepsilon_t - \theta_L \varepsilon_{t-L} - \theta_{2L} \varepsilon_{t-2L} - \theta_{3L} \varepsilon_{t-3L} - \ldots - \theta_{QL} \varepsilon_{t-QL} \qquad \textbf{(3.16)}$$

In practice, most series exhibit both seasonal and successive period relationships. Box and Jenkins (1976) derive a complex combination of the seasonal and successive schemes that yields what is called a multiplicative ARMA model of order (p, q) X (P, Q). Essentially, the model combines terms from equations (3.14) - (3.16) and allows for successive AR and MA relations in the previous cycle by adjusting the coefficients in a multiplicative fashion that is analogous to the correction for seasonal effects by the multiplication of a seasonal index in equation (3.11) of Winters' method. Having already revealed seasonal patterns in the loss rate data, the seasonal ARMA model of order (p, q) X (P, Q) shown in equation (3.17) was chosen as the appropriate ARMA model.

$$
\begin{aligned}
X_t &= \phi_1 X_{t-1} + (\phi_L X_{t-L} + \phi_1 \phi_L X_{t-L-1}) + \dots + \phi_p X_{t-p} + (\phi_{PL} X_{t-PL} + \phi_p \phi_{PL} X_{t-PL-1}) + \\
&\quad \varepsilon_t - \theta_1 \varepsilon_{t-1} - (\theta_L \varepsilon_{t-L} - \theta_1 \theta_L \varepsilon_{t-L-1}) - \dots - \theta_q \varepsilon_{t-q} - (\theta_{QL} \varepsilon_{t-QL} - \theta_q \theta_{QL} \varepsilon_{t-QL-1}) \qquad \textbf{(3.17)}
\end{aligned}
$$

The most troublesome aspect of the smoothing techniques is determining the values for the smoothing constants. For the ARMA model, maximum likelihood estimates may be obtained for the AR and MA coefficients. As a result, the most troublesome aspect of ARMA modeling is determining the proper order (p, q) X (P, Q). For this data set, a restriction on the ARMA model's order was immediately imposed. Recalling the TSDT, the 48th MIS series has only 24 months of data values. This restricted the ARMA model to seasonal order values of P = 1 and Q = 1, where forecasts requiring data values from 12 and 13 periods earlier (seasonal and successive relations with the last cycle) were calculable. If either P = 2 or Q = 2, forecasts would require data values from 12, 13, 24 and 25 periods in the past. Since the last time series, MIS 48, had only 24 data values these forecasts would not be calculable.

Box and Jenkins (1976) proposed a method of model selection that stressed the principle of parsimony and was heavily based on trial and error. Parsimony encourages the use of the simplest model with the least number of parameters to estimate. Trial and error calls for an arbitrary model selection, parameter estimation and forecast error analysis. If the resulting forecast errors, or residuals, are random and without pattern then the model may be judged adequate. Following this procedure, a seasonal ARMA model of order (1, 1) x (1, 1) was first hypothesized and judged sufficient with respect to residual analysis for all but a few MIS time series. The greatest troubles occurred in the 46th and 47th MIS (due to 4 year enlistees) series, and to a lesser extent in the 34th and 35th MIS (due to 3 year enlistees) series. These series were all affected by manpower policies (Early-Out Programs) allowing soldiers to depart service prior to their true End of Term of Service (ETS). These policies were not available at all times throughout the analysis and validation data years, but when they were they had the greatest effect in the summer months which traditionally have the highest accession, and hence the highest ETS activity. Having more data values than the 46th and 47th MIS series, the 36th and 37th MIS series managed to smooth the policy differences and produce smaller errors in validation. Since, the 46th and 47th MIS series had only one year of data with large loss rates in May, June and July, these large rates were carried forward into all future years, even if such policies were not in effect. As a result, large errors were observed each of these future months, thus creating a clear pattern in the residuals. To eliminate this effect, the large rates needed to be smoothed or down weighted by surrounding data. Accordingly, a model of order (1, 3)X(1, 1) was hypothesized and evaluated. The model sufficiently corrected the deficiencies of the previous model, although not entirely. The

greatest limitation was still the number of data values in the later MIS series and additional data was not readily available.

In summary, a ARMA model of order (1,3)x(1,1) was judged appropriate for this data set and adopted. The model selection is subject to the limitations of this data set, but is still useful to demonstrate the method and compare its performance with the others.

Another feature of ARMA modeling is worthy of mention. Since ARMA relations simply express future values as a function of past ones, other variables may be incorporated into the regression-like relation. The coeficients of these variables may then be estimated using linear least squares methods to achieve maximum liklihood estimates, (Makridakis and Wheelwright, 1978). For example, factors such as the absence or presence of certain manpower policies, or econometric variables effecting soldiers' decisions to leave service, could be included to affect forecasts. Currently, such effects are achieved by user manipulation of the the exponential smoothing constant to obtain anticipated results. As such, ARMA models offer a more analytically sound and documentable approach. Due to the limited size of the data set, this aspect of ARMA modeling was not explored. However, a discussion of the effect of policies and world events on the observed errors during validation is contained in Chapter IV.

D. ANALYSIS OF FORECAST ERRORS

To evaluate each forecast method with respect to the others, it was necessary to derive some measures and displays that quantify and summarize the error in forecast. Since ELIM's fundamental purpose is to model enlisted force strength, the error analysis is strength based with an actual or known strength in time t and MIS s equal to the number of soldiers

entering MIS s at the beginning of time t. Strengths constructed from known initial values and a series of forecasted loss rates are used as the basis for error calculations. Implicitly, this convention captures the cumulative effect of errors since current strengths are always calculated from the last estimated strength and forecasted loss rate. This approach proved more relevant and useful that simply comparing estimated and forecasted loss rates since these comparisons fail to capture the cumulative effect of the errors and provide little intuitive appeal as to the meaning of results. Additionally, since loss rates range between zero and one, loss rate comparisons are not computationally well behaved when aggregate and summarized.

The remainder of this section provides the details of the strength based analysis of forecast errors. The measures of effectiveness subsection covers the construction of forecasted strengths and associated measures of error. The displays subsection defines and describes the displays used to analyze and summarize the errors in forecasted strength.

1. Measures of Effectiveness

a. Forecasted Strength

Recalling the validation portion of the Time Series Data template (TSDT), a forecasted strength n resides in each (t, s) cell defined by calendar months $t > (T'+1)$, and months in service $s = \{2, 3, ..., 48\}$. The cells defined by $(T'+1, s)$ are known force strengths calculable from the last data strengths and loss rates contained in cells (T', s). The cells defined by (t, 1), $1 \leq t \leq 144$, are also known strengths determined by the total number of soldiers accesses into the active Army during calender month t.[8] Together, these known

[8] Since this thesis never forecasts into a truly unknown future, these accession totals are known. In reality, ELIM would use projected or target accession numbers for each month into the planning horizon.

strengths given by $N_0(t,s)$ serve as the initial values from which strength forecasts are obtained when estimated loss rates (λ) are applied. In total, there are 3,336 (71 months x 47 MIS's) validation cells with strengths forecasted by the relation,

$$n(t, s) = [1-\lambda(t-1, s-1)]\, n(t-1, s-1) \qquad \textbf{(3.18)}$$

where if $t = (T'+1) = 73$ or $s = 1$, then $n(t,s) = N_0(t, s)$.

By the design of the validation data set, for every forecasted strength n(t, s) the actual strength N(t, s) is known. Further, strengths may be summed across each row of the TSDT to obtain the total first term force strength for each calender month t. Actual and forecasted total are available and given by N(t) and $n(t)$ respectively. Note, such a summing down the columns of the TSDT would have no logical interpretation since during any month in real time, there is only one cohort serving in any one particular month in service.

b. Monthly Relative Error (PE) in Forecasted Strength

The actual and forecasted first term force strength totals may be calculated for every month t according to,

$$N(t)=\sum_{s=1}^{s=48} N(t,s) \qquad \text{and} \qquad \textbf{(3.19)}$$

$$n(t)=\sum_{s=1}^{s=48} n(t,s)\ . \qquad \textbf{(3.20)}$$

The total error in forecast strength may then be calculated for every month t according to,

$$E(t) = n(t) - N(t)\ . \qquad \textbf{(3.21)}$$

An error in forecasted strength is useful only with respect to the magnitude

of its component strengths and thus does not compare well to other errors. To illustrate this, an error of 10 soldiers when the actual strength is 15, is much different than an error of 10 soldiers when the actual strength is 150. To overcome this, the relative error in forecasted strength is calculated by,

$$RE(t) = \frac{n(t) - N(t)}{N(t)} . \tag{3.22}$$

A natural summary measure of error with respect to time, monthly relative errors in forecasted strength were calculated for each forecasting methodology. The resulting measures were plotted in line graphs to show relative performance across the methods.

c. *MIS Mean Relative Error (MRE)*

With respect to month in service (MIS), the error in forecasted strength is appropriately measured by the mean relative error (MRE). Since the estimated and known force strength are known for every (t, s) cell, the corresponding relative error in forecasted strength may be calculated according to

$$RE(t, s) = \frac{n(t, s) - N(t, s)}{N(t, s)} . \tag{3.23}$$

Relative errors may then be summarized by Mean Relative Errors (MRE's) given by,

$$MRE(s) = \frac{1}{(t_1 - t_0 + 1)} \sum_{t_0}^{t_1} RE(t, s) . \tag{3.24}$$

MIS MRE's were calculated for each forecasting method to compare and

contrast errors with respect to month in service. The resulting measures were plotted in line graphs to show relative performance across the methods.

d. Grand Mean Percent Error

All of the relative errors defined thus far may be converted to percentage errors simply by multiplying the quantities by a factor of 100. A grand mean percent error in forecasted strength was calculated for each forecasting method by averaging all (t, s) relative errors and multiplying by a factor of 100. The author believes this summary measure of performance conforms to the current method by which ELIM users express the model's strength based performance. While it is a grossly aggregated summary measure that obscures each method's unique behaviors, it does provide a single intuitive measure by which to compare performance.

2. Displays

a. Histograms

Histograms are useful to display the shape and distribution of data values across their range of observation. Accordingly, histograms were constructed to display and compare the distribution of each forecasting method's errors. Two types of histograms were constructed.

(1) Grand Histograms. Grand histograms were constructed for each forecasting method using all (*t, s*) relative errors in forecasted strength. These histograms provide a holistic view of each method's error distribution and readily reveal any similarities or differences. Grand histograms are located in Chapter IV.

(2) Paired (Calender Year, YOS) Histograms. Histograms were constructed for subsets of the (t,s) relative errors defined by all Calendar Year and YOS pairs within the validation portion of the TSDT. For example, a histogram for Calender Year 1989 and YOS 1 - written Hist(1989, YOS 1) - contains all relative errors defined by the intersection of Calender Year 1989 rows and YOS 1 columns in the TSDT. Figure 3.2 illustrates this idea and shows a total 24 histograms were created for each method of loss rate forecast. Located in Appendix F, these histograms reveal changes in the distribution of errors as YOS and calender year change.

	YOS 1 s = (2,..., 12)	**YOS 2** s = (13,..., 24)	**YOS 3** s = (25,..., 36)	**YOS 4** s = (37, ..., 49)
1989 t =(74, ..., 84)	Hist(1989, YOS 1)	Hist(1989, YOS 2)	Hist(1989, YOS 3)	Hist(1989, YOS 4)
1990 t=(85, ..., 96)	Hist(1990, YOS 1)	Hist(1990, YOS 2)	Hist(1990, YOS 3)	Hist(1990, YOS 4)
1991 t=(97, ..., 108)	Hist(1991, YOS 1)	Hist(1991, YOS 2)	Hist(1991, YOS 3)	Hist(1991, YOS 4)
1992 t=(109, ..., 120)	Hist(1992, YOS 1)	Hist(1992, YOS 2)	Hist(1992, YOS 3)	Hist(1992, YOS 4)
1993 t=(121, ..., 132)	Hist(1993, YOS 1)	Hist(1993, YOS 2)	Hist(1993, YOS 3)	Hist(1993, YOS 4)
1994 t=(133, ..., 148)	Hist(1994, YOS 1)	Hist(1994, YOS 2)	Hist(1994, YOS 3)	Hist(1994, YOS 4)

Figure 3.2 Representation of the Paired (Calendar Year, YOS) Histogram organization. Actual Histograms are located in Appendix F. Recalling the TSDT, the row where t=73 and the column where s = 1 each contain known strengths and hence do not have errors in forecast. For this reason, YOS 1 histograms and 1989 histograms contain twelve less error values than the others with 144 erros. The (1989, YOS1) histogram is effected by the known row and column and thus has 23 less error values.

b. Boxplots

Boxplots are a graphical display that show a measure of location (the median), a measure of dispersion (the interquartile range) and the presence of any outliers. Additionally, they indicate whether the distribution of the data is symmetric or skewed. (Rice, 1995) For these reasons and the ability to neatly arrange boxplots from each forecast method in one figure made them an attractive display for the analysis of error in forecasted strength. Like the histograms, several types of boxplots were constructed.

(1) Grand Boxplots. Grand boxplots summarizing all relative errors obtained from each forecast method were constructed and arranged side by side for comparison. Chapter IV contains the grand boxplots.

(2) Yearly Boxplots. Yearly boxplots summarize all (t, s) relative errors in forecasted strength belonging to a specific calendar year within the forecast horizon.

(3) Year of Service (YOS) Boxplots. YOS boxplots summarize all (t, s) relative errors in forecasted strength belonging to a specific year of service.

c. Line Graphs

Line graphs were constructed to display the calculated measures of effectiveness for each method simultaneously. Specifically, one line graph displays the actual and forecasted first term total force strengths with respect to real time. Another graph displays the monthly relative errors in forecasted strength (Monthly RE's), and last graph contains the mean relative errors with respect to month in service (MIS MRE's). All line graphs are contained in Chapter IV.

IV. RESULTS

A. OVERVIEW

This chapter's objective is to communicate the results and insights obtained from the analysis of each forecasting method's errors. The reader is reminded that errors are *strength based*. Accordingly, throughout this chapter the term *error* implies the *relative error in forecasted strength*. Also note, due to the cumulative nature of strength based error analysis, increasing error trends indicate a general tendency underestimate loss rates. Likewise, decreasing error trends indicate a general tendency to overestimate loss rates. These relations will become more apparent with the presentation of the results, but they are introduced here for the reader's contemplation.

First, Section B presents the distribution of errors for each forecasting method and identifies any similarities or differences across the methods. The distributions are displayed in the grand histograms and boxplots explained in Chapter III. Section C then presents each method's performance with respect to the measures of effectiveness derived in Chapter III. The measures are displayed in line graph to allows visual comparison across the methods. Section D addresses the effect of significant world events and manpower policies on the observed errors. Section E summarizes the insights gained from the results.

B. THE DISTRIBUTION OF ERRORS

1. Histograms

Figure 4.1 contains the grand histograms depicting the distribution of relative errors obtained from each forecasting method. The histograms are so similar that no

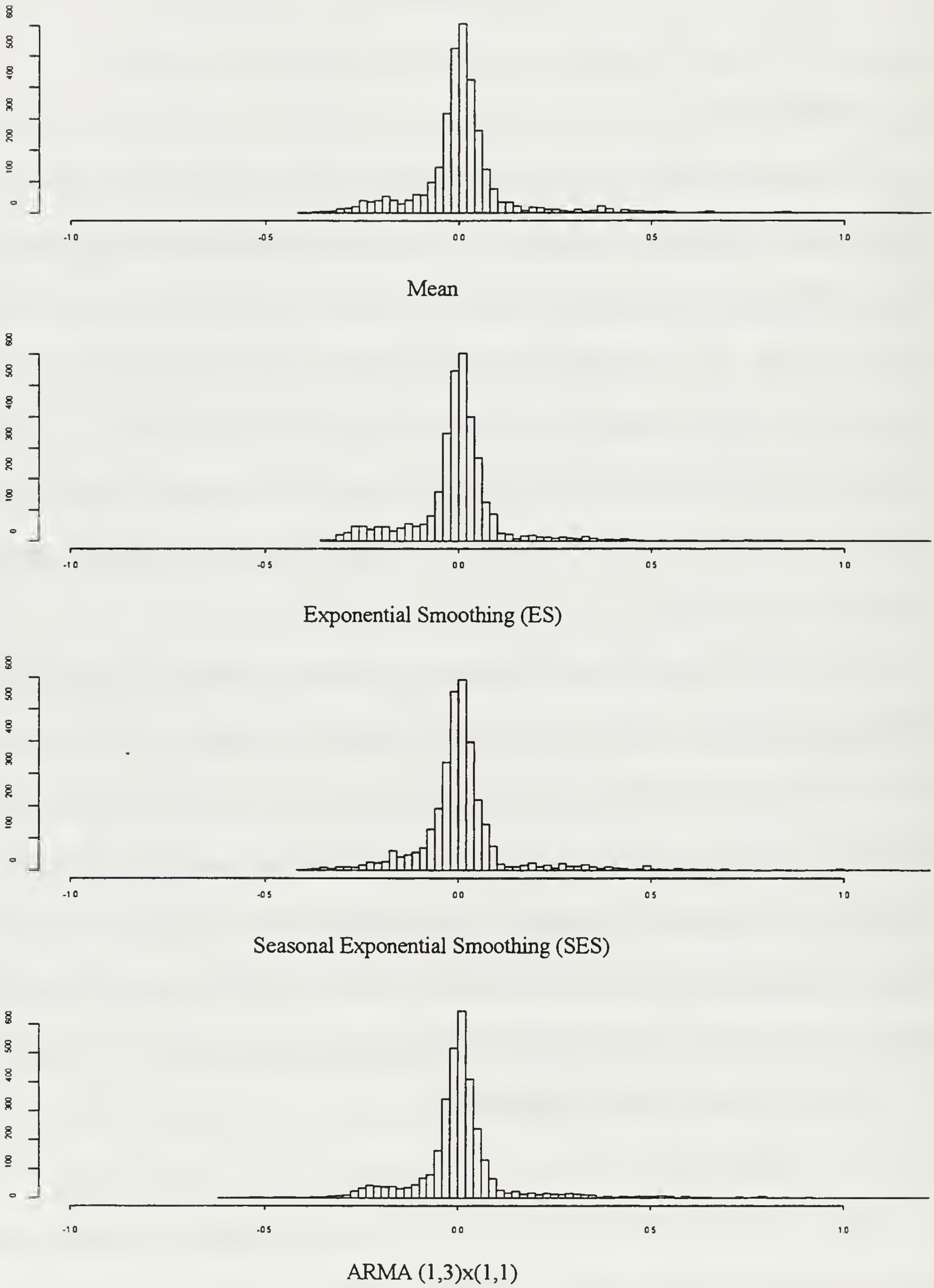

Figure 4.1 Grand histograms displaying the distribution of relative errors [RE(t,s)] for each loss rate forecasting method.

significant differences may be noted. All of the methods produce errors that are approximately normally distributed about zero relative error in forecasted strength. Close to zero, the left tails of the distributions contain slightly more observations than in the right tail, thus indicating a slightly greater tendency to underestimate the force strength. Perhaps the only notable difference between the distributions is the presence of a longer left tail in the ARMA method's histogram. This indicates the ARMA method produces the largest underestimates of force strength on occasion. This characteristic may not be entirely bad if the underestimates are viewed as counterweights to the gross overestimates which are common to all methods.

The paired (Calender Year, YOS) histograms contained in Appendix F, also show remarkably similar error distributions across the methods. Comparing each (Calender Year, YOS) histogram with its respective counterparts across the methods reveals these similarities. Also notable in these displays is the general trend toward wider error distributions as Calender Year and YOS increase. For calender year increases this trend is reasonable and obvious. It conforms to the generally accepted idea that forecasts further into the future are less reliable than those closer to the present. The trend's appearance as a function of YOS indicates that all methods produce reliable loss rate estimates for soldiers in their first and second year of service, but less reliable ones for those in their third and fourth years.

2. Boxplots

Figure 4.2 contains the grand boxplots constructed for each forecasting method. Due to the scale, the boxplots provide little information with respect to the median

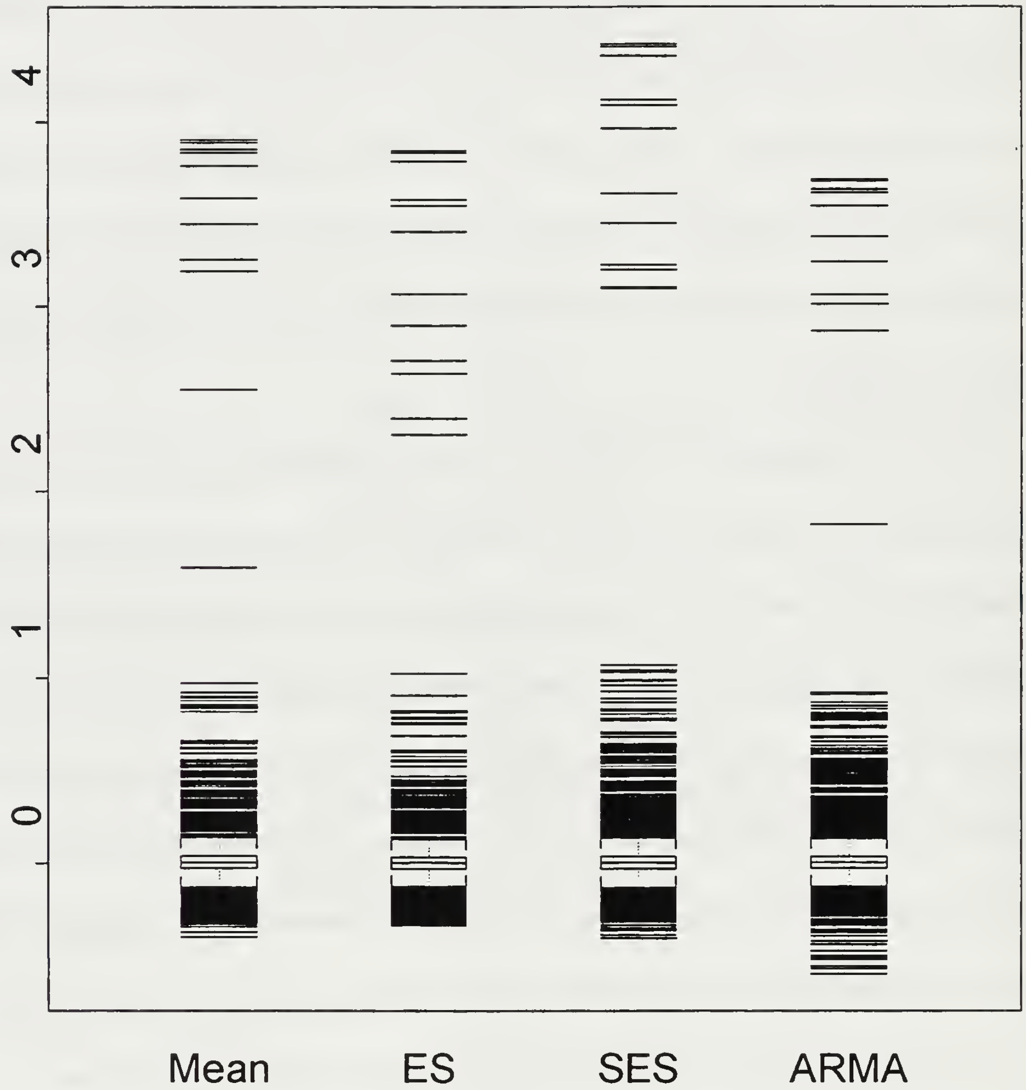

Figure 4.2 Grand boxplot of errors in forecasted strength for each loss rate forecasting method.

and interquartile range of the error distributions, but they provide great insight as to the presence of outliers. The largest positive outliers occur in all methods due to an abnormal phenomena in the data. Apparently, an overwhelming majority of soldiers belonging to cohort 9001 were 3 year term enlistees. Accordingly, most of this cohort departed service in the 36th MIS where the forecasted loss rate for all methods was still relatively small. More concretely, only 127 soldiers were in service at the beginning of their 37th MIS while 613 were forecasted. The large error in forecasted strength was then carried forward into each of the cohort's remaining months in service (37 -48).

The boxplots also highlight the small number of larger underestimates unique to the ARMA method and first identified in the histograms. Setting this difference aside, the boxplots are all relatively similar. They show each method produces errors with a median at about zero, small interquartile range center at the median and a similar distribution of outliers.

Appendix G contains boxplots constructed from each method's yearly and YOS errors. These plots confirm the trend toward wider error distributions as Calender Year and YOS increase and they provide more useful information regarding the distribution of outliers. Specifically, trends of overestimation and underestimation were deduced from these plots and reckoned with known policies and world events. The trend are discussed further in section D of this chapter and summarized in Table 4.2.

C. MEASURES OF EFFECTIVENESS

1. Estimated First Term Force Strength

Figure 4.3 is a plot of the actual and estimated first term force strength from 8901 thru 9412. Perhaps more than any other, this plot shows the performance similarities between the loss rate forecasting methods. Also notable in this plot, is the recurrent dips in first term force strength during the summer months. These dips may be explained by the fact that accessions are generally higher in the summer months, hence creating many summertime ETS losses when those soldiers reach their ETS. Since the general trend was decreasing first term force size throughout this period, there were more ETS losses (enlisted from three or four years ago) than there were new accessions during these months. Hence the sudden dips appeared Another phenomena that may also be at work during this time frame is the congressional requirement for the Army to meet its obligated Force Strength Allowance (FSA) by the end

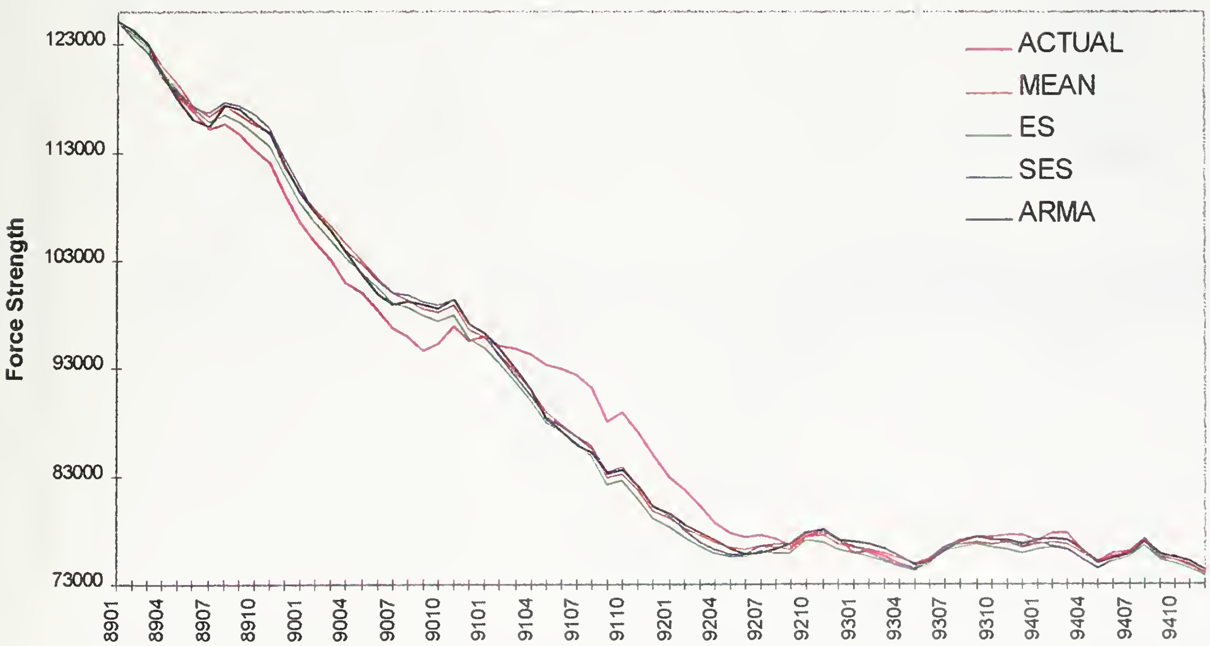

Figure 4.3 Actual and forecasted total first term force strength.

of each fiscal year. A responsive means of achieving this goal is by controlling accessions and early releases according to need. This phenomena may explain the reversed spike in September 1994. Appendix H contains plots of the accession totals between 8301 and 9412 which contributed to these observations.

2. Monthly Relative Errors in Estimated Strength

Figure 4.4 displays the monthly relative errors in forecasted strength for each method The plot allows for greater scrutiny of the differences between the methods, but fails to indicate one that is clearly superior. As was seen in Figure 4 3, all methods follow the same general error patterns Notable on this plot is the ARMA methods accentuation of the summertime dips. This behavior was discussed in detail during the development of the ARMA model in Chapter III and led to the adoption of the (1,3)x(1,1) order to minimize the effect.

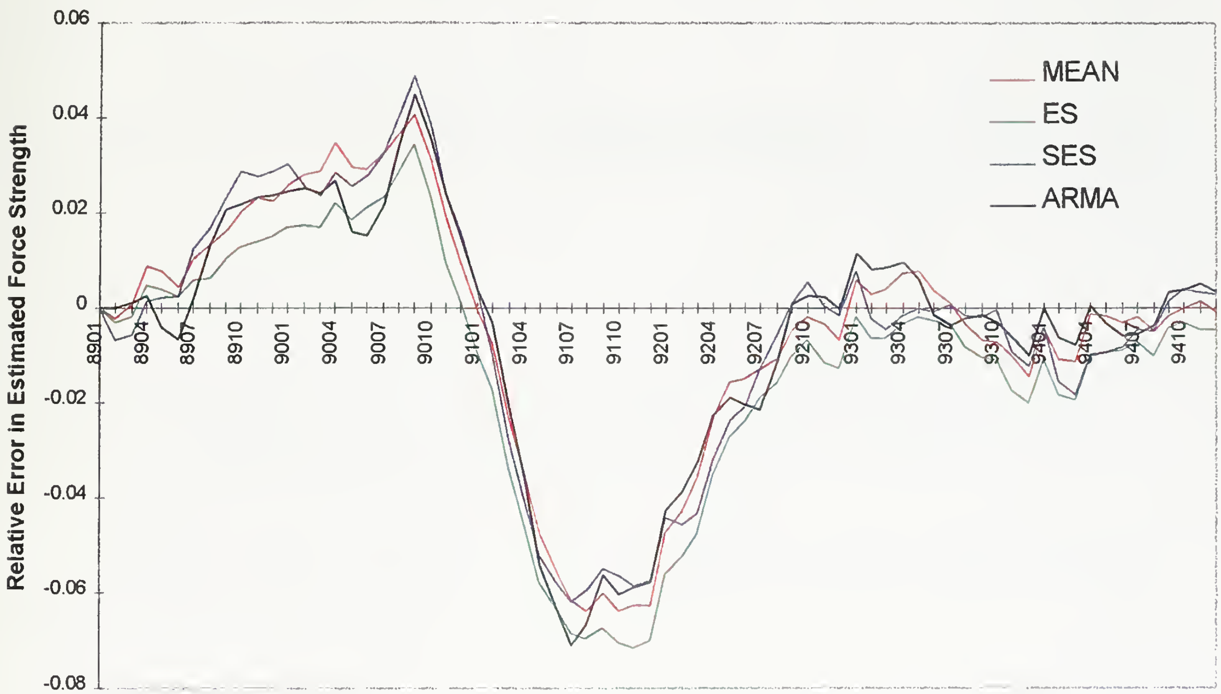

Figure 4.4 Monthly relative error in forecasted strength.

3. MIS Mean Relative Errors

Figure 4.5 contains a plot of each method's Mean Relative Error (MRE) with respect to Month in Service (MIS). The plot shows remarkably similar performance between the methods out to the 30th MIS. Following that time, all methods experience great and varied errors. The mean and ARMA forecast methods show sharp drops in the 47th and 48th MIS's due to the cumulative nature of the strength based errors and their larger overestimates in the 42d - 46th MIS's. While this plot identifies some clear differences in performance, it fails to identify a superior method.

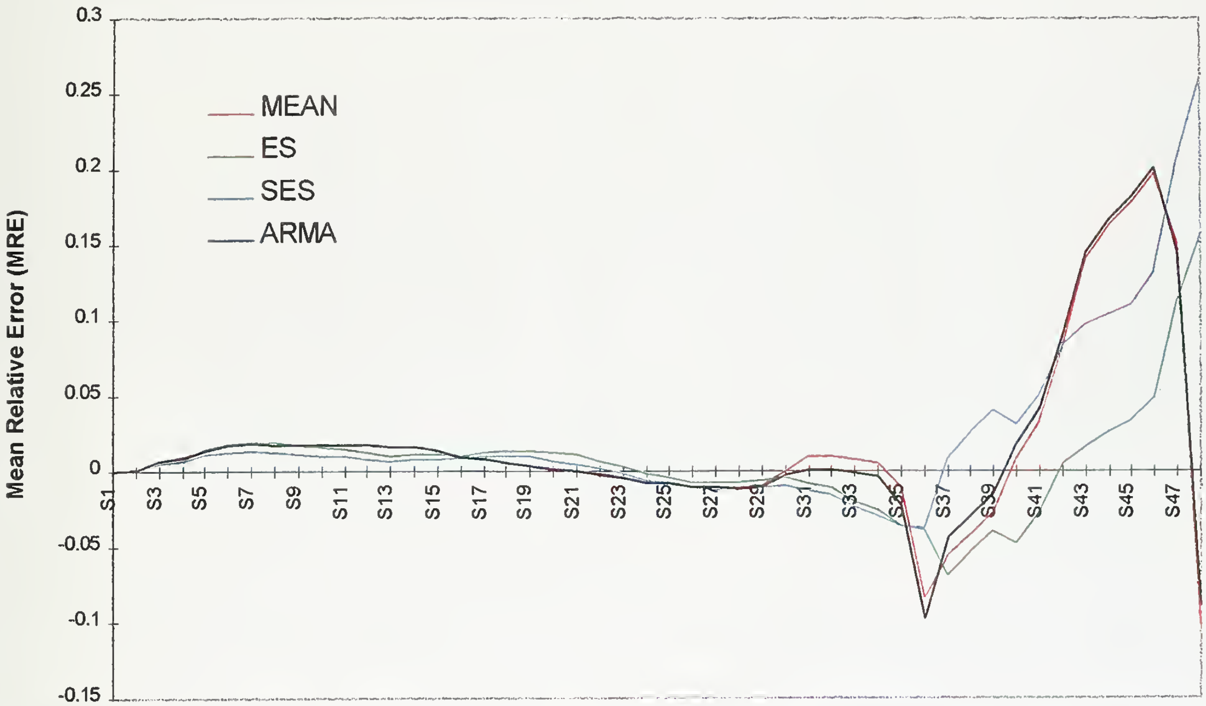

Figure 4.5 Mean relative error (MRE) in forecasted strength for each MIS series.

4. Grand Mean Percent Errors

The grand mean percent errors (MPE's) in forecasted strength for each method are contained in Table 4.1. The measure provides a single summary measure of performance with respect to actual observations but, because of its gross aggregation, it may not be used to

decisively identify any one method as more accurate than the rest. An interesting aspect of this statistic is the seasonal exponential smoothing's poor performance. This is most likely due to the strong seasonal behavior demonstrated in the data, but not during the validation years. A similar problem was observed in the seasonal ARMA model, but its effects was lessened by increasing the order of the moving average component from one to three. Such an adjustment to the seasonal exponential smoothing model could only be accomplished through manual manipulation of the smoothing constants. This offers little in the way of analytical rigor and hence was not performed.

Method	Grand MPE
Mean	1.83%
Exponential Smoothing	0.55%
Seasonal Exponential Smoothing	2.40%
ARMA (1,3)x(1,1)	1.84%

Table 4.1 Grand Mean Percent Errors

D. POLICY EFFECTS ON THE RESULTS

All of the results presented thus far indicate some peculiar behavior with respect to calender years - particularly around 1990 and 1991. This error is explained by the occurrence of Operations Desert Shield and Desert Storm during those years. During the Desert Shield buildup and throughout the war, the U.S. Army instituted a "stop-loss" that prevented soldiers from leaving service for routine reasons (most End of Term of Service (ETS) separations were not allowed). Since future wars are rarely predicted with accuracy, the forecast models

were not adjusted for this policy factor. In fact, only the ARMA model could analytically incorporate such external factors into the prediction of loss rates. The other models must be subjectively manipulated to achieve such effects.[9]

The impact of Desert Shield and Desert Storm manpower policies may be seen in many of the measures and displays presented, but nowhere quite as clearly and understandably as in Figure 4.4. Prior to the war, the general trend was toward overestimation of the first term force strength. This trend was then reversed by the stop-loss. During the war and while the stop-loss was in effect loss rates were grossly overestimated and hence strength was underestimated. Following the war and after the stop-loss was repealed, loss rates were underestimated as all those who should have left service due to ETS were now allowed to do so. This reversed the strength trend, causing largely underestimated force strength to close toward zero. By December 1991, the force strength predictions had not fully recovered from the war but were beginning to climb. Many of the peculiar aspect of the results may be explained by policies and events. In Figure 4.3 the MRE for the later MIS's were erratic and large. This may be explained by the absence or presence of early-out programs which allow soldiers to leave service prior to the actual ETS for a variety of reasons. These programs help manage the force size and meet end strength goals. Obviously, all early-out programs were halted by the stop-loss instituted during Operations Desert Shield and Storm.

Table 4.2 summarizes the significant events, manpower policies and resulting estimationtrends that effected forecasts between 1989 and 1994. The table indicates general

[9] Policy effects may be incorporated into the smoothing techniques only by adjusting the alpha and/or gamma smoothing constants to achieve reasonably expected results. Additionally, forecasted loss rates may be manually adjusted up or down to incorporate the effects of known or planned policies and events.

			General Trend in forecasted ...	
Dates	**Significant Events**	**Manpower Policies and Trends**	**Loss Rates**	**Strengths**
Jan 89 - Sep 90	Berlin Wall falls in November 1989.	Early-out programs in effect. Force stabilization programs following Reagan buildup - trend toward reduction.	Underestimated	Overestimated with an increasing gap between forecasted and actual strength.
Aug 90 - Dec 90	Iraq invades Kuwait (Aug 90). Desert Shield (Sep 90 - Dec 91).	Stop-Loss Initiated with Desert Shield.	Overestimated	Overestimated with a decreasing gap between forecasted and actual strength.
Jan 91 - Jun 91	Desert Storm (Jan 91 - Mar 91). Major redeployments of US forces (Mar 91 - Jun 91).	Stop-loss in effect.	Overestimated	Underestimated with an increasing gap between actual and forecasted strength
Jul 91 - Dec 92	Continued presence in Persian Gulf and Northern Iraq (Kurds) (Jun 91 - Dec 91).	Stop loss repealed mid-year allowing all those held in service to depart. Pride in service effects may have reduced loss behavior with respect to certain types of loss.	Underestimated	Underestimated with no clear trend - halted the widening gap between actual and estimated strength
Jan 92 - Dec 92	Bottom-Up review resulting in force reductions and realignments	Aggressive early-out programs initiated. BRAC and European Force Reductions.	Underestimated	Underestimated with a decreasing gap between forecasted and actual strength
Jan 93 - Dec 93	No events with significant manpower effects	Force reductions and realignments.	Underestimated	Slightly overestimated to slightly underestimated
Jan 94- Dec 94	No events with significant manpower effects	Force reductions and realignment, approaching stability.	No Clear Trend	No Clear Trend

Table 4.2 The effect of significant events and manpower policies and trends on forecasted loss rates and strengths. The information in this table is a synthesis of all the results of this thesis, news paper clippings chronicling Desert Shield/Storm events from the Jacksonville Daily News, Jacksonville North Carolina, and policy information provided by ODCSPER.

trends in forecasted loss rates and force strength that may be seen throughout the results and displays found in this chapter and the appendices. The trends are a synthesis of all the results with a heavy reliance on the histograms and boxplots contained in Appendices F and G.

E. SUMMARY

The following list summarizes the insights gained from the time series analysis of U.S. Army loss rates and the resulting errors in forecasted strength.

1. The error distributions obtained from each forecasting do not diifer in any uniform way.

2. For all methods, forecasts further into the future are less reliable than those closer to the present.

3. For all methods, forecasts for soldier's in their first and second year of service are more reliable than for those in their third and fourth year of service.

4. No one loss rate forecasting method provides consistently more accurate estimates of strength with respect to time or month in service.

6. All forecasting methods react similarly to significant events and manpower policies.

V. CONCLUSION

A. SUMMARY

The analysis and prediction of personnel loss behavior is critical to effective manpower planning and to the U.S. Army's ELIM-COMPLIP system. As such, monthly historical loss rates were constructed from personnel loss/gain event records. These rates represent the proportion of soldiers from each cohort under study that left the Army during each month of service in their first term enlistment. The study included only those soldiers belonging to C-Group 1, and only while serving in there first term. The results of the study must always be caveated by this C-group 1 restriction, but they remain valid and important since over 45% of the Army's total accessions during the study period were C-Group 1 soldiers.

Monthly loss rate data from January 1983 to December 1988 was organized into a times series data template from which future loss rates were forecasted. The time series analysis techniques all sought to identify patterns in the data and forecast them into the future via time based extrapolations. The forecasting methods explored were the arithmetic mean, exponential smoothing, seasonal exponential smoothing, and an autoregressive moving average model.

Forecasted loss rates were used to construct monthly first term force strength projections six years beyond the last data month - from January 1989 to December 1994. The forecasts were then compared to known force strengths for the same periods. Comparisons were quantified and summarized by relative errors in forecasted strength. The errors were displayed in a variety of forms to allow performance comparisons between the loss rate forecasting methods.

The analysis of error in forecasted strength revealed no significant performance differences between the loss rate forecasting methods. The methods' error distributions were remarkably similar and all methods performed similarly with respect to world events and policies that affected first term force strength. In terms of complexity and sophistication, the methods rank from simplest to most complex according to; mean, exponential smoothing, seasonal exponential smoothing, autoregressive moving average. In terms of the mean percent error in forecasted strength, the methods rank in the order of best to worst according to; exponential smoothing (0.55%), mean (1.83%), autoregressive moving average (1.84%), seasonal exponential smoothing (2.40%). While these mean percent errors are useful and contribute to the overall evaluation, they obscure the unique behavior of each method and may not be used to definitively identify any one method as superior to another.

B. OVERALL CONCLUSIONS

Since no significant differences in performance were noted, the simplest methods may be viewed as more economical, and thus favored. Accordingly, the exponential smoothing method currently employed in the ELIM-COMPLIP system has been validated as appropriate with respect to the other time series analysis method explored. Another interesting result is the viability of the arithmetic mean as an estimate of loss. Simple, understandable and effective, the arithmetic mean of past loss rates proved itself a valuable forecast that could facilitate timely answers to many manpower planning problems.

C. RECOMMENDATIONS FOR FURTHER STUDY

Capable of extension beyond the scope of this thesis, the autoregressive moving average method is worthy of further study. Able to analytically incorporate other variables

such as the absence or presence of strength affecting policies or econometric indicators of loss behavior, the model may achieve greater accuracy than demonstrated here. Such a study will require several more years of data than was available for this thesis, and an accurate record of strength affecting events and policies. Any future study of the autoregressive moving average method should also consider examination of lifetime regression and survival analysis techniques as they are also capable of incorporating the effects of other variables into forecasts.

APPENDIX A. CGD & SLC PROGRAM (SAS)

The Characteristic Group Designator and Service Life Calculator (CGD & SLC) processes the ELIM-COMPLIP raw database called the Small Tracking File (STF). The program partitions the STF into the Characteristic Groups defined by Table 2.1, and calculates service lifetimes from individual Gain/Loss records. The program was coded for SAS version 6.07 and executed on an Amdahl 5995-700A Mainframe Computer in MVS batch mode.

CGD & SLC Program Listing;:

```
//CGDSLC JOB USER=S2706,CLASS=H
//*MAIN LINES=(50)
// EXEC SASBIG
//SASIN DD DISP=SHR,DSN=MSS.S2706.STF
//SASOUT DD DISP=(OLD,KEEP),DSN=MSS.S2706.CGDSLC
//SYSIN DD *
*-------------------------------------------------------------
PROGRAM NAME: CHARACTERISTIC GROUP DESIGNATOR AND SERVICE LIFE CALCULATOR ( CGDSLC SAS)
DESCRIPTION : TRANSFORMSTHE STF TO A PERMANENT SAS DATA SET, PARTITIONED BY C-GROUP,
              WITH END-FIRST TERM SERVICE LIFE AND OVERALL SERVICE LIFE CALCULATIONS.

DATE         :  15 JUL 96
PROGRAMMER   :  CAPT E.T. DEWALD USMC
*-------------------------------------------------------------;
OPTIONS MEMSIZE=20M;
DATA SASOUT.CGDSLC;
    ATTRIB
       MSVFL     FORMAT =  3.  LABEL = 'MONTHS OF SERVICE TO FIRST LOSS'
       FL_CEN    FORMAT =  1.  LABEL = 'FIRST LOSS CENSOR INDICATOR'
       MSV_ET1   FORMAT =  3.  LABEL = 'MONTHS OF SERVICE TO END TERM 1'
       ET1_CEN   FORMAT =  1.  LABEL = 'END TERM ONE CENSOR'
       COHORT    FORMAT = $4.  LABEL = 'COHORT YYMM'
       C_GROUP   FORMAT = $1.  LABEL = 'CHARACTERISTIC GROUP'
       AFQT      FORMAT =  2.  LABEL = 'ARMED FORCES QUAL TEST SCORE'
       MENT_CAT  FORMAT = $2.  LABEL = 'MENTAL CATEGORY'
       RACE      FORMAT =  1.  LABEL = 'RACE'
       TERM      FORMAT =  1.  LABEL = 'INITIAL TERM OF SERVICE'
       CIVED     FORMAT = $1.  LABEL = 'CIVILIAN EDUCATION CODE'
       ED_CAT    FORMAT = $3.  LABEL = 'EDUCATIONAL CATEGORY'
       AGEENTRY  FORMAT =  3.  LABEL = 'AGE AT ENTRY TO SERVICE'
       VEL_FLAG  FORMAT = $1.  LABEL = 'VARIABLE ENLISTMENT INDICATOR'
       CURR_TT   FORMAT =  2.  LABEL = 'CURRENT TRAINING TIME';
    SET SASIN.STF2(READ=SEMPERFI);
    IF GENDER = 'F' THEN DELETE;
    IF COMPONT ^= 'R' THEN DELETE;
    /* MENTAL CATEGORY DERIVATION FROM AFQT */
    IF (AFQT <= 20) THEN MENT_CAT = '5 ';
       ELSE IF (AFQT >  20) AND (AFQT <= 30) THEN MENT_CAT = '4 ';
       ELSE IF (AFQT >  30) AND (AFQT <  50) THEN MENT_CAT = '3B';
       ELSE IF (AFQT >= 50) AND (AFQT <  65) THEN MENT_CAT = '3A';
       ELSE IF (AFQT >= 65) AND (AFQT <  94) THEN MENT_CAT = '2 ';
```

```
   ELSE IF (AFQT >= 94) AND (AFQT <= 98) THEN MENT_CAT = '1 ';
   ELSE DO;
        AFQT     = .;
        MENT_CAT = .;
   END;
IF TERM = 9 THEN TERM = .;

/* ED_CAT DERIVATION FROM CIVED (BRUTT-FORCE METHOD) */
IF CIVED = '0' | CIVED = '1' | CIVED = '2' | CIVED = '3' |
   CIVED = '4' | CIVED = '5' | CIVED = '6' | CIVED = '7' |
   CIVED = '8' | CIVED = 'A' | CIVED = 'B' | CIVED = 'C' |
   CIVED = 'D' | CIVED = 'W'  THEN ED_CAT = 'NHD';
   ELSE IF CIVED = 'H' | CIVED = 'I' | CIVED = 'J' | CIVED = 'K' |
           CIVED = 'L' | CIVED = 'M' | CIVED = 'N' | CIVED = 'O' |
           CIVED = 'P' | CIVED = 'Q' | CIVED = 'R' | CIVED = 'S' |
           CIVED = 'T' | CIVED = 'U' | CIVED = 'V' | CIVED = 'Y'
           THEN ED_CAT = 'HDP';
   ELSE IF CIVED = 'E' THEN ED_CAT = 'HSD';
   ELSE IF CIVED = 'F' | CIVED = 'G' THEN ED_CAT = 'GED';
   ELSE ED_CAT = .;
IF AGEENTRY = 999 THEN AGEENTRY = .;
AGEENTRY = INT(AGEENTRY/12}; /*CONVERT MONTHS TO YEARS*/
/* POLICY BASED TRANSFORMATION OF VEL_FLAG BASED ON VEL PROGRAM */
IF COHORT < '8504'  THEN VEL_FLAG = 'N';
   ELSE IF TERM = 2 THEN VEL_FLAG = 'V';
   ELSE IF TERM = 5 THEN VEL_FLAG = 'N';
   ELSE IF TERM = 6 THEN VEL_FLAG = 'N';
   ELSE IF VEL_FLAG = ' ' THEN VEL_FLAG = .;
   ELSE VEL_FLAG = VEL_FLAG;
/* CURR_TT ADJUSTMENT BASED ON PAGE 2.10 OF GRC ELIM EXECUTIVE */
/* OVERVIEW BRIEFING DATED 950201.  NOTE: CUR_TT IS CHARCTER   */
/* DATA AND CURR_TT IS NUMERIC.                                */
IF (CUR_TT > 13) AND (CUR_TT ^= 99) THEN CURR_TT = 13;
   ELSE IF CUR_TT = 99 THEN CURR_TT = .;
   ELSE IF CUR_TT < 2 THEN CURR_TT = 2;
   ELSE CURR_TT = CUR_TT;
 /* CHARACTERISTIC GROUP DERIVATION */
IF (((ED_CAT = 'HSD') OR (ED_CAT = 'HDP') OR (ED_CAT = 'GED')) AND
   ((MENT_CAT = '1 ') OR (MENT_CAT = '2 ') OR (MENT_CAT = '3A')) AND
   ((TERM = 3) OR (TERM = 4))) THEN C_GROUP ='1';
IF (((ED_CAT = 'HSD') OR (ED_CAT = 'HDP') OR (ED_CAT = 'GED')) AND
    (MENT_CAT = '3B') AND
   ((TERM = 3) OR (TERM = 4))) THEN C_GROUP ='2';
IF (((ED_CAT = 'HSD') OR (ED_CAT = 'HDP') OR (ED_CAT = 'GED')) AND
   ((MENT_CAT = '4 ') OR (MENT_CAT = '5 ')) AND
   ((TERM = 3) OR (TERM = 4))) THEN C_GROUP = '3';
IF ((ED_CAT = 'NHD') AND
   ((MENT_CAT = '1 ') OR (MENT_CAT = '2 ') OR (MENT_CAT = '3A')) AND
   ((TERM = 3) OR (TERM = 4))) THEN C_GROUP = '4';
IF ((ED_CAT = 'NHD') AND
   ((MENT_CAT = '3B') OR (MENT_CAT = '4 '} OR (MENT_CAT = '5 ')} AND
   ((TERM = 3) OR (TERM = 4})) THEN C_GROUP = '5';
IF ((TERM = 2) OR (TERM = 5) OR (TERM = 6)) THEN C_GROUP = '9';
ARRAY MSVS{24} MSV1 - MSV24;
ARRAY EVENTS{24} EVENT1 - EVENT24;
ARRAY LGRE{24} $;
/* TRANSFORM EVENTS INTO LOSS, GAIN, RE-ENLIST/EXTEND (LGRE} */
DO INDEX = 1 TO 24;
   IF EVENTS{INDEX} = 'NPG' | EVENTS{INDEX} = 'RMC' |
      EVENTS{INDEX} = 'L90' | EVENTS{INDEX} = 'G90' |
      EVENTS{INDEX} = 'NPA' | EVENTS{INDEX} = 'RSV' |
      EVENTS{INDEX} = 'OTG' THEN LGRE{INDEX} = 'GAIN';
   ELSE IF EVENTS{INDEX} = 'EDP' | EVENTS{INDEX} = 'EMP' |
           EVENTS{INDEX} = 'ERL' | EVENTS{INDEX} = 'DFR' |
           EVENTS{INDEX} = 'BLK' | EVENTS{INDEX} = 'ETS' |
           EVENTS{INDEX} = 'MCD' | EVENTS{INDEX} = 'HRD' |
           EVENTS{INDEX} = 'LLL' | EVENTS{INDEX} = 'MPP' |
           EVENTS{INDEX} = 'OSR' | EVENTS{INDEX} = 'OTH' |
           EVENTS{INDEX} = 'PHY' | EVENTS{INDEX} = 'RET' |
           EVENTS{INDEX} = 'SCH' | EVENTS{INDEX} = 'TDP' |
           EVENTS{INDEX} = 'UFT' THEN LGRE{INDEX} = 'LOSS';
```

```
      ELSE IF EVENTS{INDEX} = 'IMR' | EVENTS{INDEX} = 'EXT' THEN
              LGRE{INDEX} = 'EXRE';
      ELSE LGRE{INDEX} = .;
   END;

   /* DERIVING MONTHS OF SERVICE TO FIRST LOSS FROM EVENT LIST DATA */
   MSVFL = 0;
   FL_CEN = 0; /* NOT CENSORED */
   FL_FLAG = 'F';
   DO INDEX = 1 TO NEVENTS;
      IF ((FL_FLAG = 'F') AND (LGRE{INDEX} = 'LOSS')) THEN DO;
         MSVFL = MSVS{INDEX};
         FL_FLAG = 'T';
      END;
   END;
   IF FL_FLAG = 'F' THEN DO
      FL_CEN = 1; /* CENSORED */
      YY1 = INT(COHORT/100);
      MM1 = COHORT - (YY1*100);
      YY2 = INT(CEN_DATE/100);
      MM2 = CEN_DATE - (YY2*100);
      MSVFL = (12-MM1) + MM2 + ((YY2-(YY1+1))*12);
   END;
   /* DERIVING MONTHS OF SERVICE TO END 1ST TERM (MSV_ET1) */
   MSVEXRE = 0;
   EXRE_FLG = 'F';
   DO INDEX = 1 TO NEVENTS;
      IF ((EXRE_FLG = 'F') AND (LGRE{INDEX} = 'EXRE')) THEN DO;
         MSVEXRE = MSVS{INDEX};
         EXRE_FLG = 'T';
      END;
   END;
   IF EXRE_FLG = 'F' THEN DO
      YY1 = INT(COHORT/100);
      MM1 = COHORT - (YY1*100);
      YY2 = INT(CEN_DATE/100);
      MM2 = CEN_DATE - (YY2*100);
      MSVEXRE = (12-MM1) + MM2 + ((YY2-(YY1+1))*12);
   END;
   IF MSVFL < MSVEXRE THEN DO;
      MSV_ET1 = MSVFL;
      ET1_CEN = 0;
   END;
   IF MSVEXRE < MSVFL THEN DO;
      MSV_ET1 = MSVEXRE;
      ET1_CEN = 0;
   END;
   IF MSVEXRE = MSVFL THEN DO;
      MSV_ET1 = MSVEXRE;
      ET1_CEN = 1;
   END;
   /*CORRECTING SMALL PERCENT (.1) OF UNREASONABLE DATA*/
   IF ((MSV_ET1) > (TERM * 12)) THEN DO;
      IF VEL_FLAG = 'V' THEN MSV_ET1 = (CURR_TT + (TERM*12));
      ELSE MSV_ET1 = (TERM*12);
   END;
   DROP SSN GENDER CIVED BP_ENTDT ETS_DATE COMPONT O_BASD C_BASD
        CUR_TT TRAILOST NEVENTS MSV1 - MSV24 EVENT1 - EVENT24
        LGRE1 - LGRE24 CEN_DATE FL_FLAG YY1 MM1 YY2 MM2 INDEX
        MSVEXRE EXRE_FLG;
RUN;
```

APPENDIX B. TSDG PROGRAM (PASCAL)

The Time Series Data Generator (TSDG) processes a homogeneous subpopulation data set created by the CGD & SLC program listed in Appendix A. The TSDG creates a historical time series data set conforming to the data template described in Chapter III, Section A. The program was coded in Borland's Turbo Pascal Version 1.5 for Windows 3.1 and executed on a 486/66 PC computer.

<u>TSDG Listing</u>:

```
program TimeSeriesDataGenerator;
{***********************************************************************
{ FileName    : tsdg.pas
{ Date        : 16 Aug 96
{ Programmer  : Edward T. DeWald
{***********************************************************************}
  uses WinCrt, WinDos;
const FIRSTYEAR =  83;
      LASTYEAR  =  95;
      RMAX      = 156; { ((LASTYEAR-FIRSTYEAR)+1)*12 }
      TMAX      = 156;
      SMAXIMUM  =  72; {SMAXIMUMIMUM is max possible months in service for
                        soldier of interest - CG9 allows 6 year terms,
                        and 6 x 12 = 72}

{NOTE: T => Real Time Indexing,
       S => Service Time Indexing,
       R => Cohort Time Indexing,
       and the relation T = S + R - 1 holds}

type DataRecordType   = record
                           LossCount   : Integer;
                           AtRiskCount : Integer;
                        end;
     DataMatrixType   = array[1..RMAX, 1..SMAXIMUM] of DataRecordType;
     InFileType       = record
                           Name : String[26];
                           SMAX : Integer;
                        end;
     InFileArrayType  = array[1..8] of InFileType;

var  DataMatrix                                 : DataMatrixType;
     InFileSpec                                 : InFileArrayType;
     InFile,NRFile,NLFile,LRFile                : Text;
     RIndex, TIndex, SIndex                     : Integer;
     Cohort,Time,LifeTime,Censor,YY,MM,SMAX     : Integer;
     Hour, Minute, Second, Sec100               : Word;
     InPath, OutPath                            : string[20];
     FileNumber                                 : Integer;
     TString, SString                           : string[3];

begin
    {Initialization I/O}
    clrscr;
    writeln('Time Series Data Generator');
```

```
writeln('        TSDG.PAS');
writeln('  Capt E.T. DeWald USMC');
writeln('==========================');
writeln;
writeln;
GetTime(Hour, Minute, Second, Sec100);
writeln('Start Time: ', Hour, ':', Minute, ':', Second);
writeln;
{Init Files}
InPath  := 'd:\datain\' ;
OutPath := 'd:\dataout\';
InFileSpec[1].Name := 'cg1'   ;   InFileSpec[1].SMAX := 48;
InFileSpec[2].Name := 'cg13'  ;   InFileSpec[2].SMAX := 36;
InFileSpec[3].Name := 'cg14'  ;   InFileSpec[3].SMAX := 48;
InFileSpec[4].Name := 'cg2'   ;   InFileSpec[4].SMAX := 48;
InFileSpec[5].Name := 'cg3'   ;   InFileSpec[5].SMAX := 48;
InFileSpec[6].Name := 'cg4'   ;   InFileSpec[6].SMAX := 48;
InFileSpec[7].Name := 'cg5'   ;   InFileSpec[7].SMAX := 48;
InFileSpec[8].Name := 'cg9'   ;   InFileSpec[8].SMAX := 72;
{Main Algorithm]
for FileNumber := 1 to 8 do begin
  {Init Current Files:  NR->Num at Risk File, NL->NumLossEventsFile,
                         1->RxS structure   ,  2->RxT structure.}
  assign(Infile,  InPath+InFileSpec[FileNumber].Name);
  assign(NRFile, OutPath+InFileSpec[FileNumber].Name+'nr.txt');
  assign(NLFile, OutPath+InFileSpec[FileNumber].Name+'nl.txt');
  assign(LRFile, OutPath+InFileSpec[FileNumber].Name+'lr.txt');
  reset(InFile);
  rewrite(NRFile);
  rewrite(NLFile);
  rewrite(LRFile);
  {Init DataMatrix}
  for RIndex := 1 to RMAX do begin
      for SIndex := 1 to SMAXIMUM do begin
          DataMatrix[RIndex,SIndex].LossCount   := 0;
          DataMatrix[RIndex,SIndex].AtRiskCount := 0;
      end;
  end;
  SMAX := InFileSpec[FileNumber].SMAX;
  {Loading Input...Counting AtRisks and Loss Events in R x S Stucture}
  write('Processing Input: ',InPath+InFileSpec[FileNumber].Name,' ...');
  repeat
     readln(InFile, Cohort, LifeTime, Censor);
     YY := Cohort div 100;
     MM := Cohort - (YY*100);
     RIndex := (YY - FIRSTYEAR)*12 + MM;
     for SIndex := 1 to LifeTime do begin
         DataMatrix[RIndex,SIndex].AtRiskCount :=
                    DataMatrix[RIndex,SIndex].AtRiskCount +1;
     end;
     {* Key Code:Count only actual loss events-not Censored lifetimes *}
     if (Censor = 0) then begin {Not Censored => Actually Ended Term1}
         DataMatrix[RIndex,SIndex].LossCount :=
                DataMatrix[RIndex,SIndex].LossCount + 1;
     end; {* End Key Code *}
  until SeekEof(InFile);
  writeln('DONE.');
  write('    writing TxS files...');
  {writing T x S Headers}
  write(NRFile, 'YYMM':10);
  write(NLFile, 'YYMM':10);
  write(LRFile, 'YYMM':10);
  for SIndex := 1 to SMAX do begin
      str(SIndex, SString);
      write(NRFile, 'S'+SString:10);
      write(NLFile, 'S'+SString:10);
      write(LRFile, 'S'+SString:15);
  end;
  writeln(NRFile);
  writeln(NLFile);
  writeln(LRFile);
```

```
      {Output Data in T x S structure}
      for TIndex := 1 to TMAX do begin
          YY := FIRSTYEAR + TIndex div 12;
          MM := TIndex mod 12;
          if MM = 0 then begin
             YY := YY - 1;
             MM := 12;
          end;
          Time := (YY*100)+MM;
          write(NRFile, Time:10);
          write(NLFile, Time:10);
          write(LRFile, Time:10);
          for SIndex := 1 to SMAX do begin
              RIndex := TIndex - SIndex + 1;
              if ((RIndex > 0) and (RIndex <= RMAX)) then begin
                 if DataMatrix[RIndex,SIndex].AtRiskCount = 0 then begin
                    write(NRFile, 'NA':10); { none at risk }
                    write(NLFile, 'NA':10);
                    write(LRFile, 'NA':15);
                 end
                 else begin
                    write(NLFile, DataMatrix[RIndex,SIndex].LossCount:10);
                    write(NRFile, DataMatrix[RIndex,SIndex].AtRiskCount:10);
                    write(LRFile, (DataMatrix[RIndex,SIndex].LossCount /
                                  DataMatrix[RIndex,SIndex].AtRiskCount):15:12);
                 end;
              end
              else begin
                   write(NRFile, 'NA':10); { none at risk }
                   write(NLFile, 'NA':10);
                   write(LRFile, 'NA':15);
              end;
          end;
          writeln(NLFile);
          writeln(NRFile);
          writeln(LRFile);
      end;
      writeln('DONE.');
      close(InFile);
      close(NRFile);
      close(NLFile);
      close(LRFile);
    end;
    {Ending Program}
    GetTime(Hour, Minute, Second, Sec100);
    writeln('End Time: ', Hour, ':', Minute, ':', Second);
    writeln;
    writeln('SEMPER FIDELIS!!');
end.
```

APPENDIX C. EXPONENTIAL SMOOTHING FUNCTIONS

The following S-PLUS functions and commands forecasts loss rates from a homogeneous data set, for each month in service, using simple Exponential Smoothing. The data set must be in the time series data template described in Chapter III, Section A2. Exponential smoothing constant (alpha's) are chosen to minimize the mean square error of forecast on the analysis data set, for each month in service time series. The smoothing constant's accuracy and range is determined by the user defined `alpha.vector`. These functions were coded in S-PLUS for Windows 3.1, Version 3.3 and executed on a 486/66 PC computer.

<u>Simple Exponential Smoothing of a Single Time Series</u>:

```
> f.exp.sm
function(x.cts, alpha)
{
        x.cts <- as.vector(x.cts[x.cts != "NA"])
        n <- length(x.cts)
        forecasts <- numeric(length = n + 1)
        forcasts[1] <- x.cts[1]
        for(index in 2:n + 1) {
                forcasts[index] <- forcasts[index - 1] + (alpha * (x.cts[
                        index - 1] - forcasts[index - 1]))
        }
        errors <- forcasts[1:n] - x.cts
        mean.error <- mean(errors)
        sigma.error <- sqrt(var(errors))
        MSE <- sum((errors^2))/(n - 1)
        forecast <- forcasts[n + 1]
        return(errors, mean.error, sigma.error, MSE, forecast)
}
```

<u>Minimum Mean Square Error Exponential Smoothing on the Times Series Data Template</u>:

```
> f.exp.sm.mmse
function(x.cts, alpha.vector)
{
        MSE <- numeric(length = (length(alpha.vector)))
        mse.min <- NULL
        alpha.min <- NULL
        for(A in 1:length(alpha.vector)) {
                temp <- f.exp.sm(x.cts, alpha.vector[A])
                MSE[A] <- temp$MSE
                if(A == 1) {
                        mse.min <- MSE[A]
                        alpha.min <- alpha.vector[A]
                        min.mean.error <- temp$mean.error
                        min.mean.sigma <- temp$sigma.error
                        forcast.min <- temp$forecast
```

```
		}
		if(MSE[A] < mse.min) {
			mse.min <- MSE[A]
			alpha.min <- alpha.vector[A]
			min.mean.error <- temp$mean.error
			min.sigma.error <- temp$sigma.error
			forcast.min <- temp$forecast
		}
	}
	return(as.matrix(rbind(alpha.min, mse.min, min.mean.error,
		min.sigma.error, forcast.min), row.names = list("alpha.min",
		"mse.min", "min.mean.error", "min.sigma.error",
		"forcast.min")))
}
```

Session Commands Producing CG1 Forecasts, allowing alpha to range from 0 to 0.6 by 0.01

```
r.all.cg1.8388.expsm.mmse.0.6.01<- as.matrix(apply(cg1.8388.cts, FUN=f.exp.sm.mmse, 2,alpha.vector= seq(0,0.6,0.01)))
row.names(r.all.cg1.8388.expsm.mmse.0.6.01) <- c("alpha.min", "mse.min", "min.mean.errors", "min.sigma.error", "forcast.min")
r.fc.cg1.8388.expsm.mmse.0.6.01 <- as.vector(as.matrix(r.all.cg1.8388.expsm.mmse.0.6.01[5,])) # FORECASTS
r.alphas.cg1.8388.expsm.mmse.0.6.01 <- as.vector(as.matrix(r.all.cg1.8388.expsm.mmse.0.6.01[1,])) # ALPHAS
```

APPENDIX D. SEASONAL EXPONENTIAL SMOOTHING

The following S-PLUS functions and commands forecast loss rates for a homogeneous subpopulation, for each month of service., using Seasonal Exponential Smoothing. The data set must be in the time series template described in Chapter III, Section A2. The smoothing constants (alpha and gamma) are chosen to minimize the mean square error of forecast on the analysis data set, for each month in service time series. Smoothing constant accuracy and range is determined by the user defined `alpha.vector` and `gamma.vector`. These functions were coded in S-PLUS for Windows 3.1, Version 3.3 and executed on a 486/66 PC computer.

<u>Seasonal Exponential Smoothing on One Tme Series</u>:

```
> f.winters.exp.sm
function(x.cts, alpha, gamma, season.length)
{
        L <- season.length
        x.cts <- as.vector(x.cts[x.cts != "NA"])
        N <- length(x.cts)
        S <- numeric(N)
        error <- numeric(N)
        I.old <- rep(1, length = L)
        I.new <- rep(NA, length = L)
        forecast <- numeric(N)
        S[1] <- x.cts[1]
        I.new[1] <- 1
        error[1] <- NA
        forcast[1] <- NA
        for(Index in 2:N) {
                Period <- Index %% L
                if(Period == 0) {
                        Period <- 12
                }
                S[Index] <- alpha * (x.cts[Index]/I.old[Period]) + (1 -
                        alpha) * S[Index - 1]
                I.new[Period] <- gamma * (x.cts[Index]/S[Index]) + (1 -
                        gamma) * I.old[Period]
                forcast[Index] <- S[Index - 1] * I.old[Period]
                error[Index] <- (forcast[Index] - x.cts[Index])
                if(Period == 12) {
                        I.new <- I.new/sum(I.new)      # Renorm the I's
                        I.old <- I.new
                        I.new <- rep(NA, times = length(I.new))
                }
        }
        error <- error[error != "NA"]
        MSE <- mean(error^2)
        mean.error <- mean(error)
```

```
        sigma.error <- sqrt(var(error))
        Forecasts <- I.old * S[N]
        Seasonal.Indicies <- I.old
        return(Seasonal.Indicies, Forecasts, mean.error, sigma.error, MSE)
}
```

Minimum Mean Square Error Seasonal Exponential Smoothing on the Data Template:

```
> f.winters.exp.sm.mmse
function(x.cts, alpha.vector, gamma.vector, season.length)
{
        L <- season.length
        MSE <- matrix(data = NA, length(alpha.vector), length(gamma.vector))
        mse.min <- NULL
        alpha.min <- NULL
        gamma.min <- NULL
        for(A in 1:length(alpha.vector)) {
                for(G in 1:length(gamma.vector)) {
                        temp <- f.winters.exp.sm(x.cts, alpha.vector[A],
                                gamma.vector[G], L)
                        MSE[A, G] <- temp$MSE
                        if((A == 1) & (G == 1)) {
                                mse.min <- MSE[A, G]
                                alpha.min <- alpha.vector[A]
                                gamma.min <- gamma.vector[G]
                                mean.min <- temp$mean.error
                                sigma.min <- temp$sigma.error
                                seasonal.min <- temp$Seasonal.Indicies
                                forcast.min <- temp$Forecast
                        }
                        if(MSE[A, G] < mse.min) {
                                mse.min <- MSE[A, G]
                                alpha.min <- alpha.vector[A]
                                gamma.min <- gamma.vector[G]
                                mean.min <- temp$mean.error
                                sigma.min <- temp$sigma.error
                                seasonal.min <- temp$Seasonal.Indicies
                                forcast.min <- temp$Forecast
                        }
                }
        }
        return(alpha.min, gamma.min, mse.min, mean.min, sigma.min,
                seasonal.min, forcast.min)
}
```

Session Commands Forecasting with Alpha and Gamma between 0 and 0.6 to 0.02 Accuracy:

```
r.cgl.8388.winters<-apply(cgl.8388.cts,FUN=f.winters.exp.sm.mmse,2,
alpha.vector=seq(0,.6,0.02),gamma.vector=seq(0,.6,0.02),12)
forecasts <-NULL
for(index in 1:48){forecasts<-cbind(forecasts,r.cgl.8388.winters[[index]]$forcast.min)}
alphas <- NULL
for(index in 1:48){alphas<-cbind(alphas,r.cgl.8388.winters[[index]]$alpha.min)}
gammas <- NULL
for(index in 1:48){gammas <- cbind(gammas, r.cgl.8388.winters[[index]]$gamma.min)}
```

APPENDIX E. ARMA FUNCTION

The following S-PLUS function and session command forecasts loss rates for a homogeneous subpopulation, for each month of service, using an autoregressive moving average (ARMA) model. The data set must be in the time series template described in Chapter III, Section A2. The order of the ARMA model is specified in the function by `model.spec`. Autoregressive and moving average coefficients are maximum likelihood estimates and are derived by S-PLUS system functions. The function was coded in S-PLUS for Windows 3.1, Version 3.3 and executed on a 486/66 PC computer.

Note: An ARIMA model of order (p, 0, q)X(P,0,Q) is equivalent to an ARMA model of (p,q)X(P,Q)

ARMA Modeling on the Data Template:

```
> f.arima.analysis
function(x.cts, npms = 72, lp = 1, ld = 0, lq = 1, BP = 1, BD = 0, BQ = 1, season = 12)
{
        for.start.time <- time(x.cts[dim(x.cts)[1],])[1] + 31
        x.mat <- as.matrix(x.cts)
        forecasts <- NULL
        model.spec<-list(list(order=c(lp,ld,lq)),list(order=c(BP,BD,BQ),period= season))#
         # Specifies Model of Order (lp,ld,lq)x(BP,BD,BQ)
        for(index in 1:dim(x.cts)[2]) {
                x <- x.mat[, index]
                x <- as.vector(x[x != "NA"])
                x.mean <- mean(x) #subtracts out mean per s-plus req. for zero mean series
                x <- x - x.mean
                x.arima.mle <- arima.mle(x, model.spec)
                x.forecast <- arima.forecast(x, n = npms, model = x.arima.mle$model)
                forecasts <- cbind(forecasts, x.forecast$mean + x.mean)
        }
        return(cts(forecasts, start = for.start.time, units = "months"))
}
```

Session Command for CG1 ARMA forecasts, model order (1,1)x(1,1), Seasonality 12 months, 60 prediction months:

```
>r.arma.11.11<-f.arima.analysis(cg1.cts,npms= 60,lp=1,ld=0,lq=1,BP=1,BD=0,BQ=1,season=12)
```

APPENDIX F. ERROR HISTOGRAMS

This Appendix contains the error histograms displaying the relative error in forecasted inventory for each time series analysis method explored. Chapter III contains a detailed description and explanation of the histograms and their construction. The following table identifies the order of their presentation within this appendix.

RELATIVE ERROR IN FORECASTED INVENTORY HISTOGRAMS

Figure	Description
F.1	Mean Forecast Method, Years 1989 - 1991, YOS 1-4
F.2	Mean Forecast Method, Years 1992 - 1994, YOS 1-4
F.3	Exponential Smoothing Forecast Method, Years 1989 - 1991, YOS 1-4
F.4	Exponential Smoothing Forecast Method, Years 1992 - 1994, YOS 1-4
F.5	Seasonal Exponential Smoothing Forecast Method, Years 1989 - 1991, YOS 1-4
F.6	Seasonal Exponential Smoothing Forecast Method, Years 1992 - 1994, YOS 1-4
F.7	ARMA (1,3)x(1,1) Forecast Method, Years 1989 - 1991, YOS 1-4
F.8	ARMA (1,3)x(1,1) Forecast Method, Years 1992 - 1994, YOS 1-4

NOTE: Figures Follow on the next eight pages, one per page.

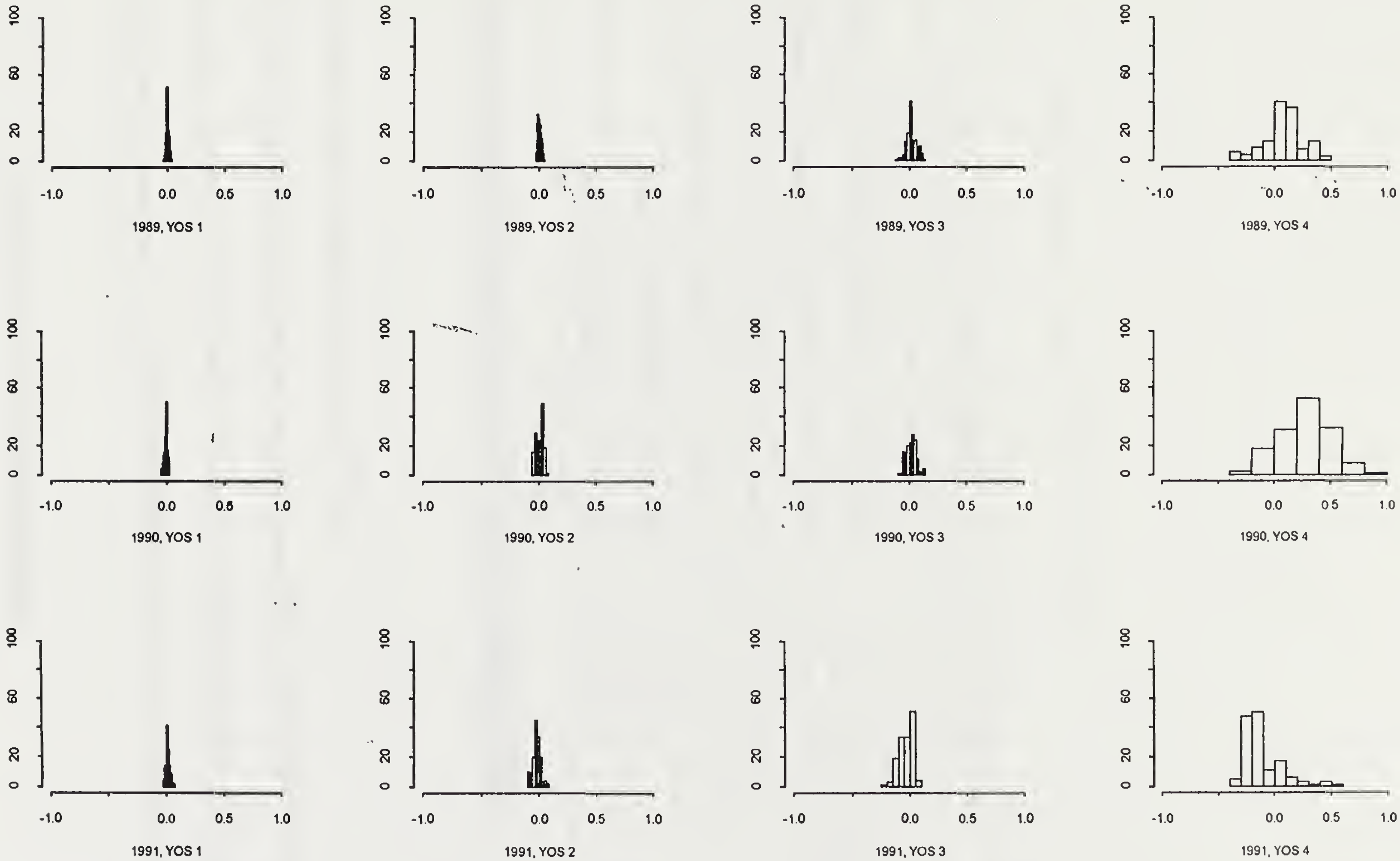

Figure F1 Relative Error in Forecasted Strength, Mean Forecast Method, 1989 - 1991, YOS 1 - 4

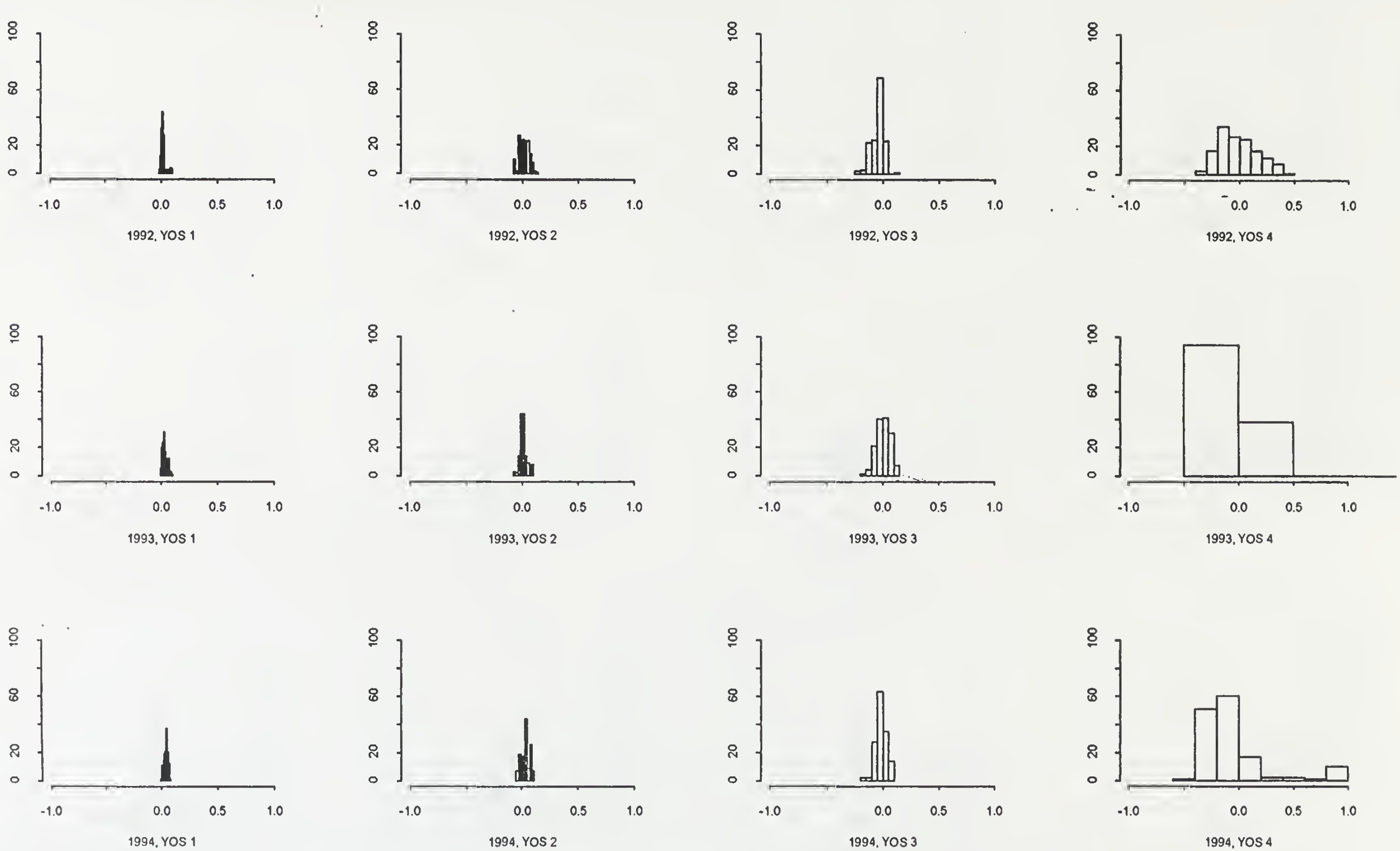

Figure F2 Relative Error in Forecasted Strength, Mean Forecast Method, 1992 - 1994, YOS 1 - 4

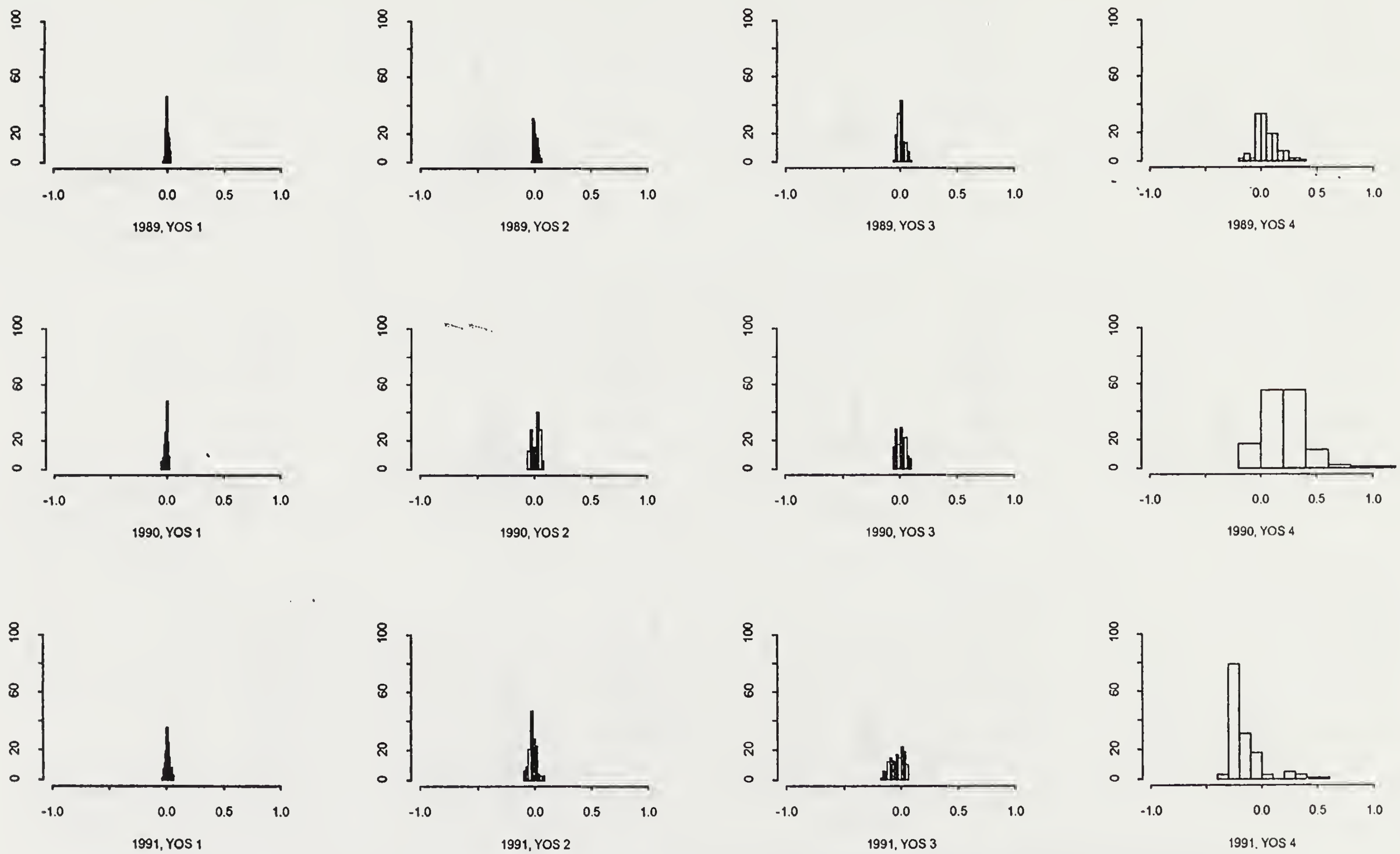

Figure F3 Relative Error in Forecasted Strength, Exponential Smoothing Forecast Method, 1989 - 1991, YOS 1 - 4

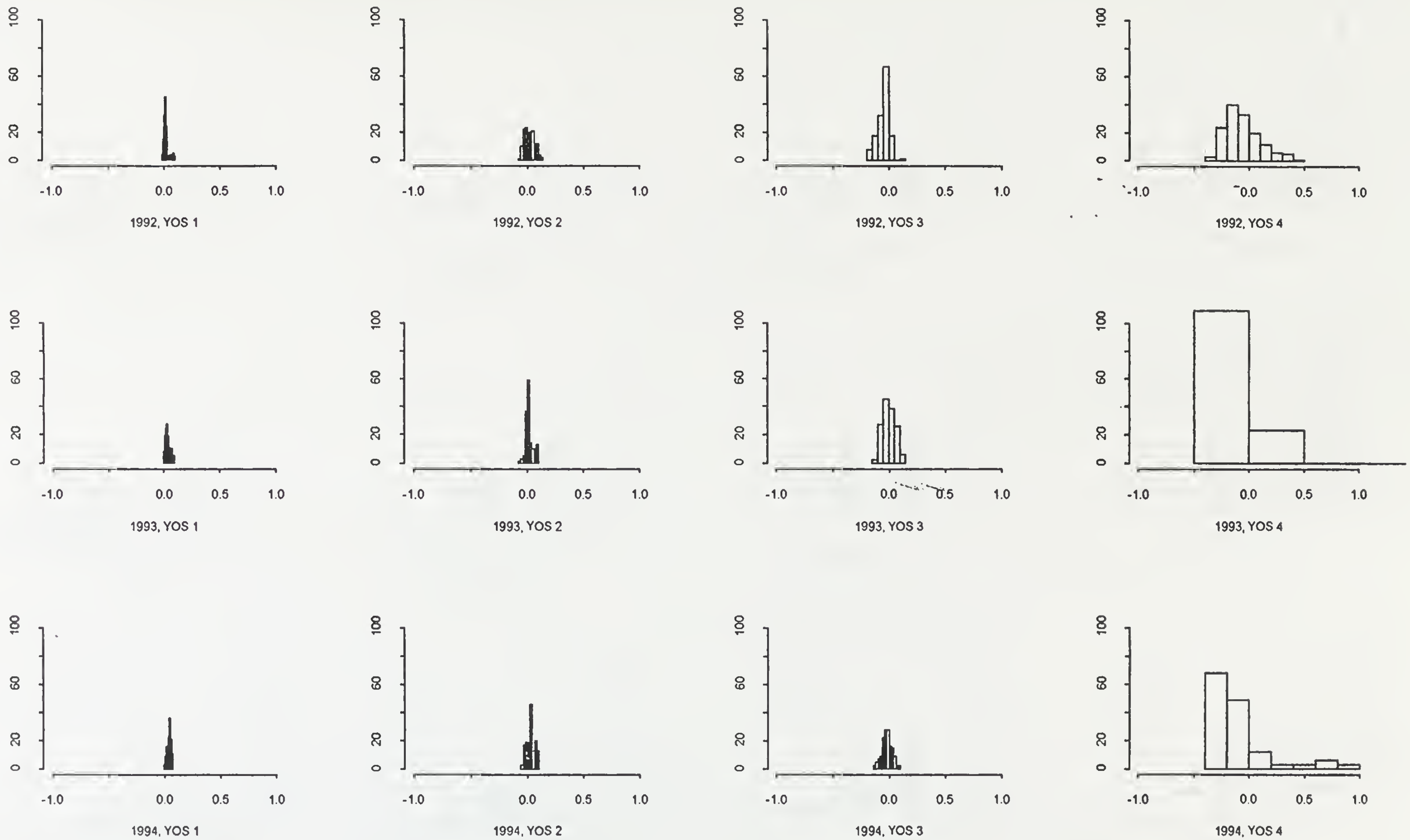

Figure F4 Relative Error in Forecasted Strength, Exponential Smoothing Forecast Method, 1992 - 1994, YOS 1 - 4

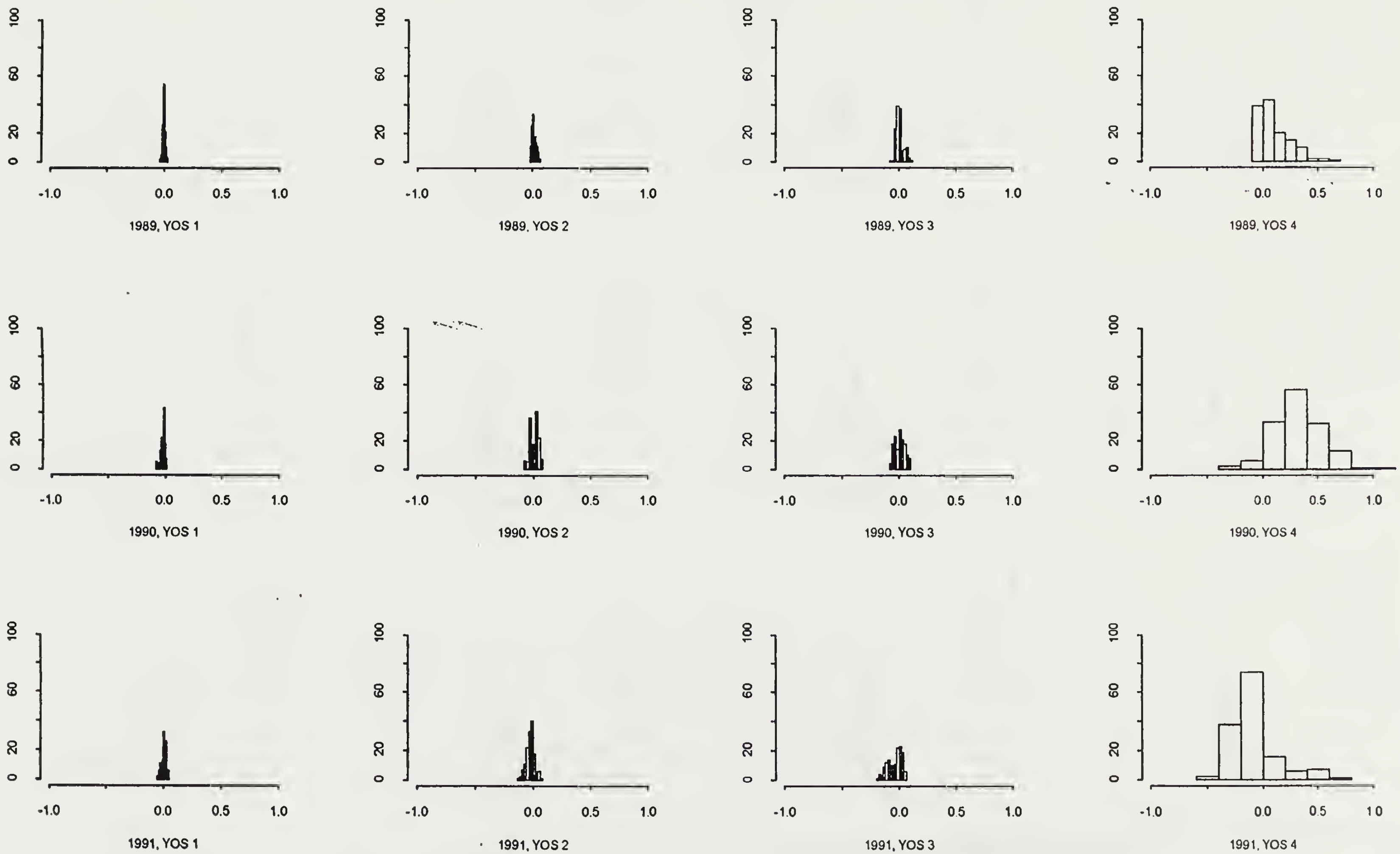

Figure F5 Relative Error in Forecasted Strength, Seasonal Exponential Smoothing Forecast Method, 1989 - 1991, YOS 1 - 4

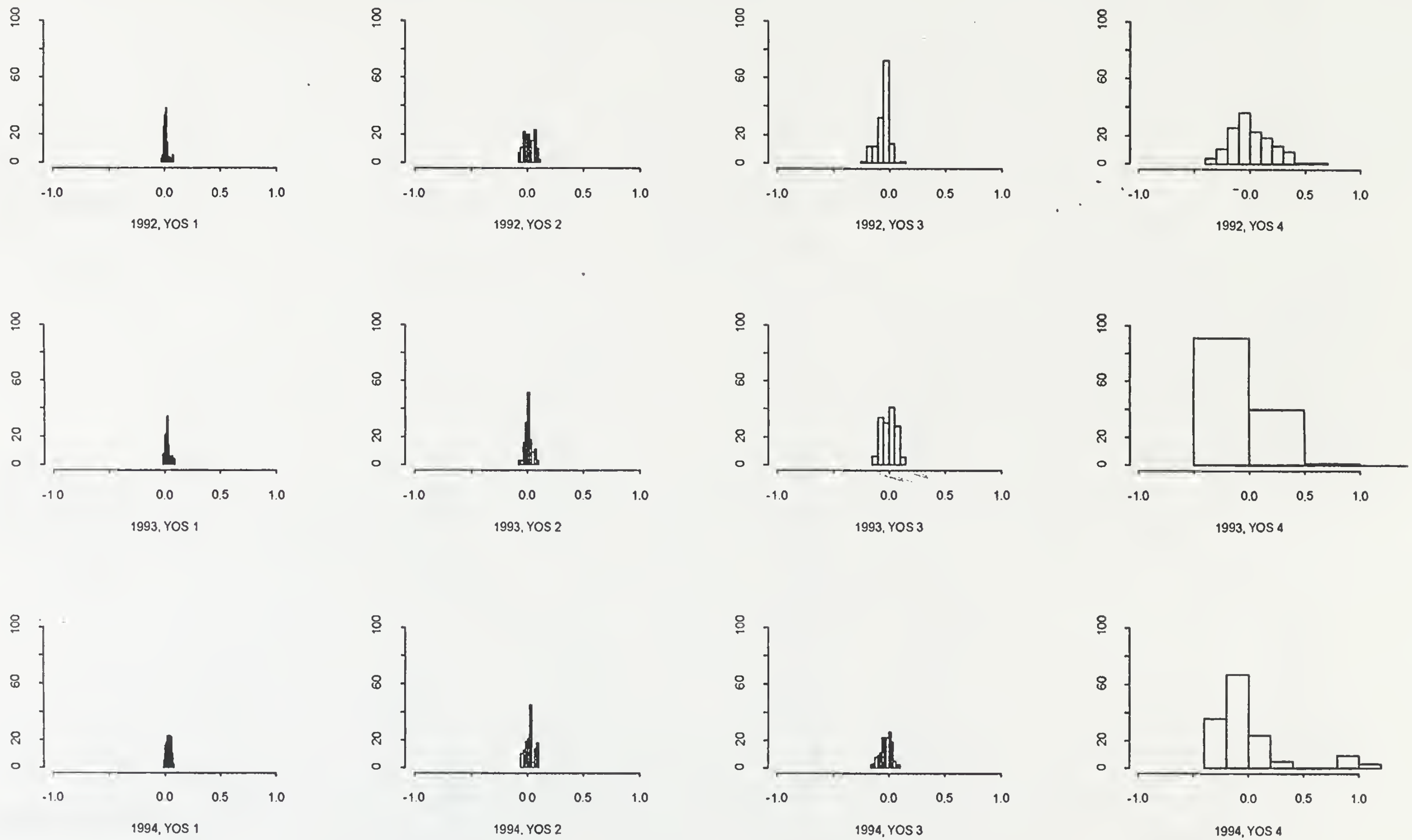

Figure F6 Relative Error in Forecasted Strength, Seasonal Exponential Smoothing Forecast Method, 1992 - 1994, YOS 1 - 4

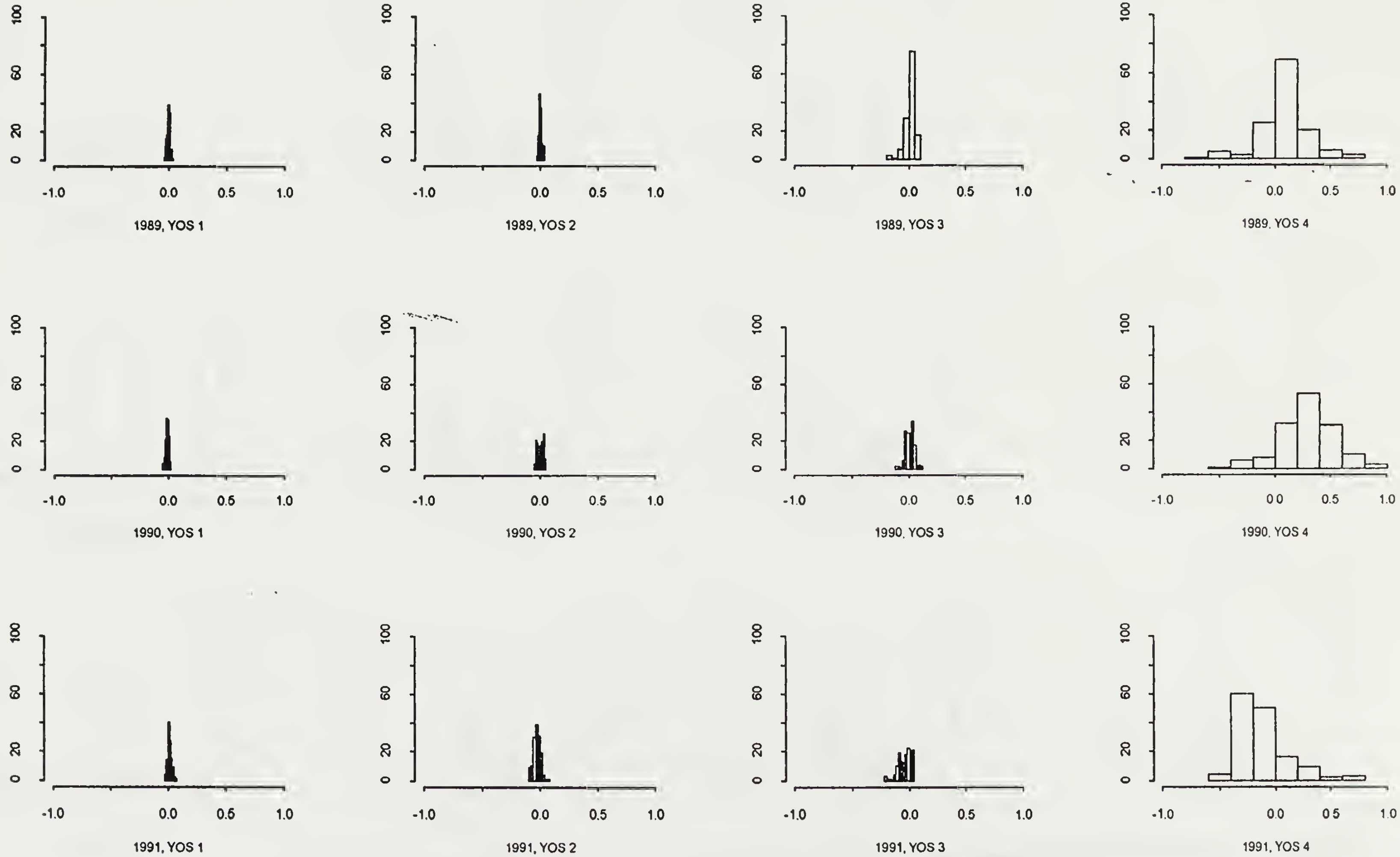

Figure F7 Relative Error in Forecasted Strength, ARMA (1,3)x(1,1) Forecast Method, 1989 - 1991, YOS 1 - 4

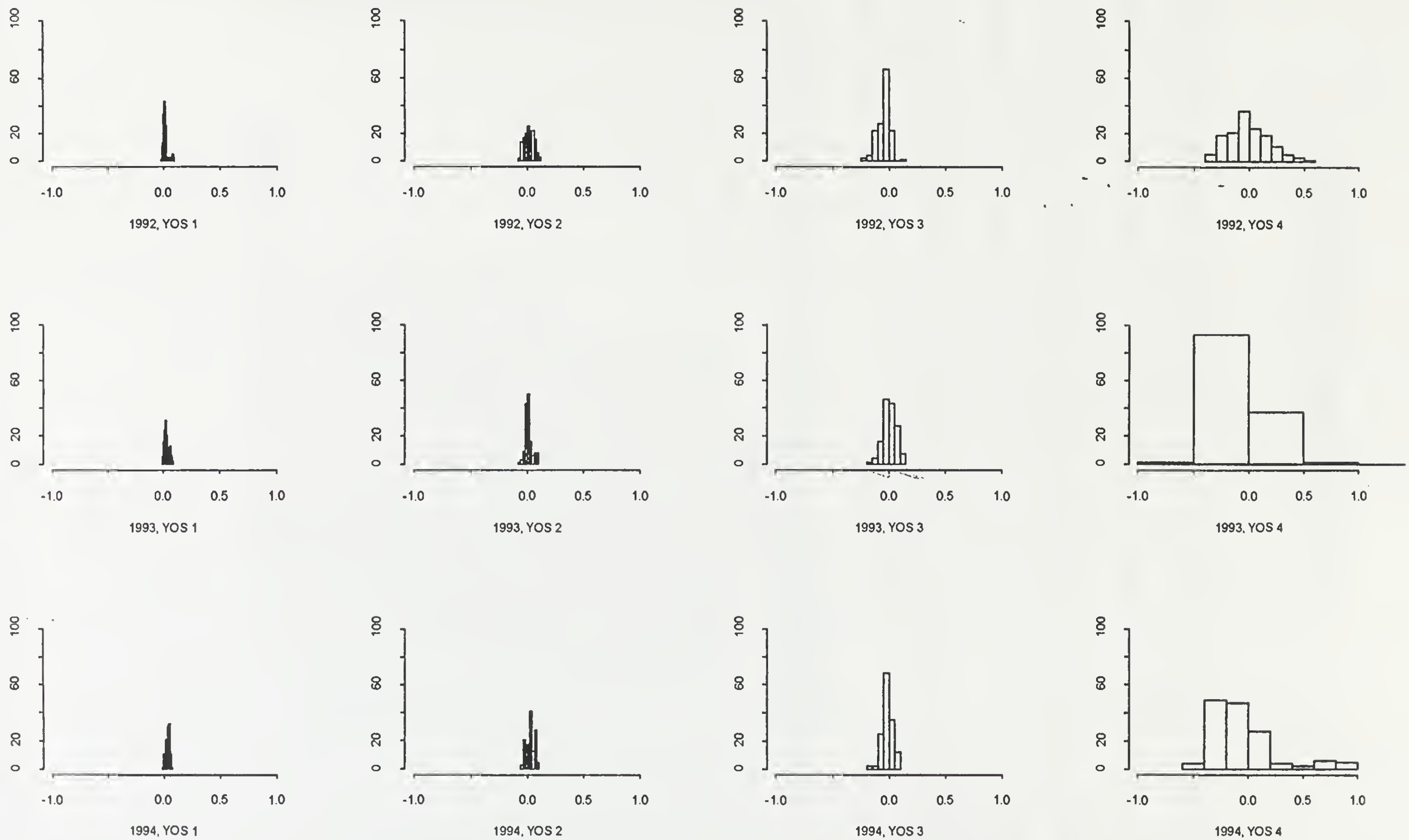

Figure F8 Relative Error in Forecasted Strength, ARMA (1,3)x(1,1) Forecast Method, 1989 - 1991, YOS 1 - 4

APPENDIX G. ERROR BOX PLOTS

This Appendix contains the error box plots displaying the relative error in forecasted inventory for Year of Service (YOS 1 - 4), and for each calendar year within the forecast horizon (1989 - 1996). In these figures, boxplots representing each of the four different forecasting methods[10] are presented side by side for direct comparison. Chapter III contains a detailed description and explanation of these boxplots and their construction. The following table identifies the order of their presentation within this appendix.

RELATIVE ERROR IN FORECASTED INVENTORY BOXPLOTS

Figure	Description
G.1	First Year of Service (YOS 1)
G.2	Second Year of Service (YOS 2)
G.3	Third Year of Service (YOS 3)
G.4	Fourth Year of Service (YOS 4)
G.5	First Year into the Forecast Horizon, 1989
G.6	Second Year into the Forecast Horizon, 1990
G.7	Third Year into the Forecast Horizon, 1991
G.8	Fourth Year into the Forecast Horizon, 1992
G.9	Fifth Year into the Forecast Horizon, 1993
G.10	Sixth Year into the Forecast Horizon, 1994

NOTE: Figures Follow on the next ten pages, one per page.

[10] Mean, Exponential Smoothing (ES), Seasonal Exponential Smoothing (SES), and AutoRegressive Moving Average (ARMA)

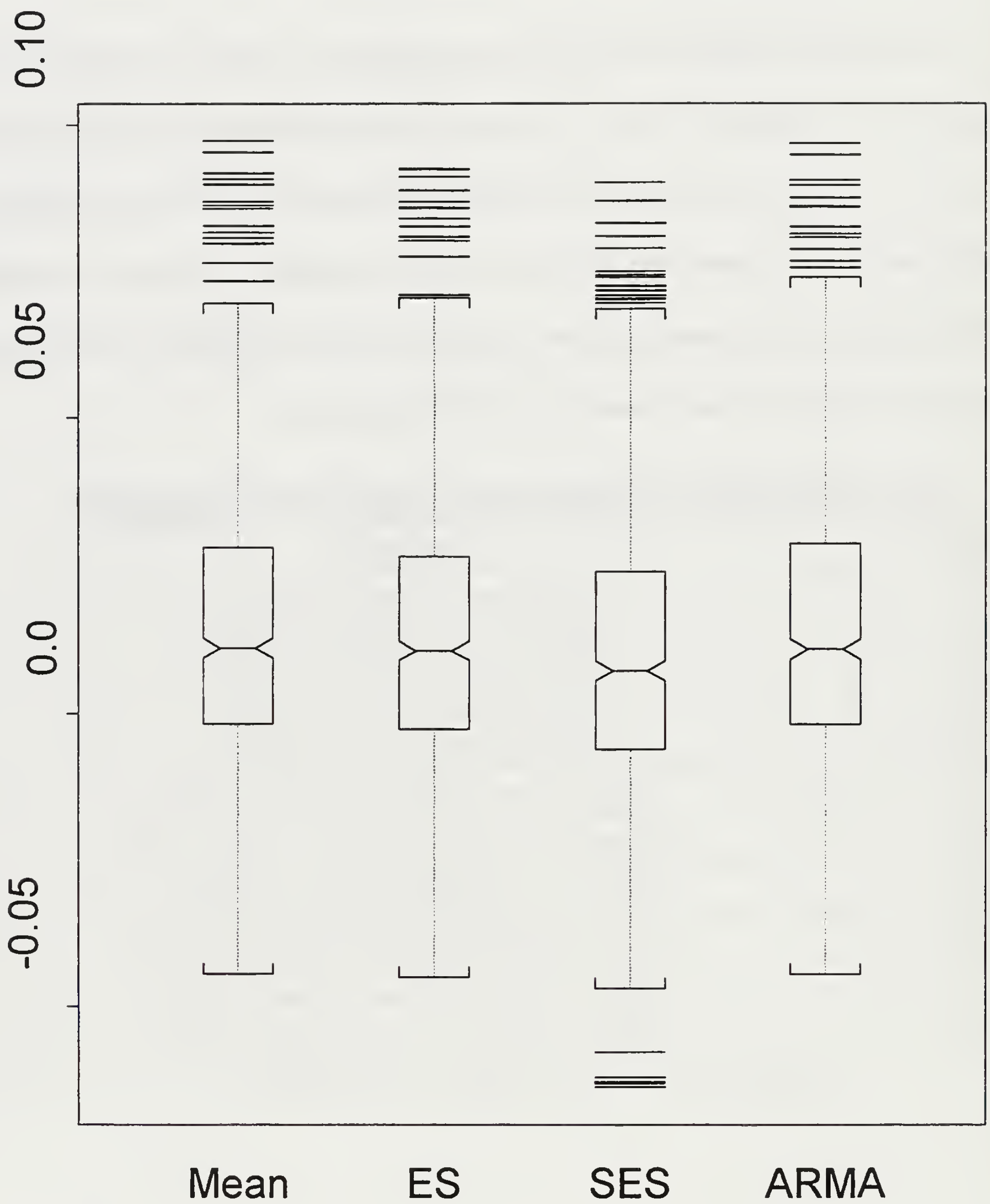

Figure G1 Relative Errors in Forecasted Strength, YOS 1.

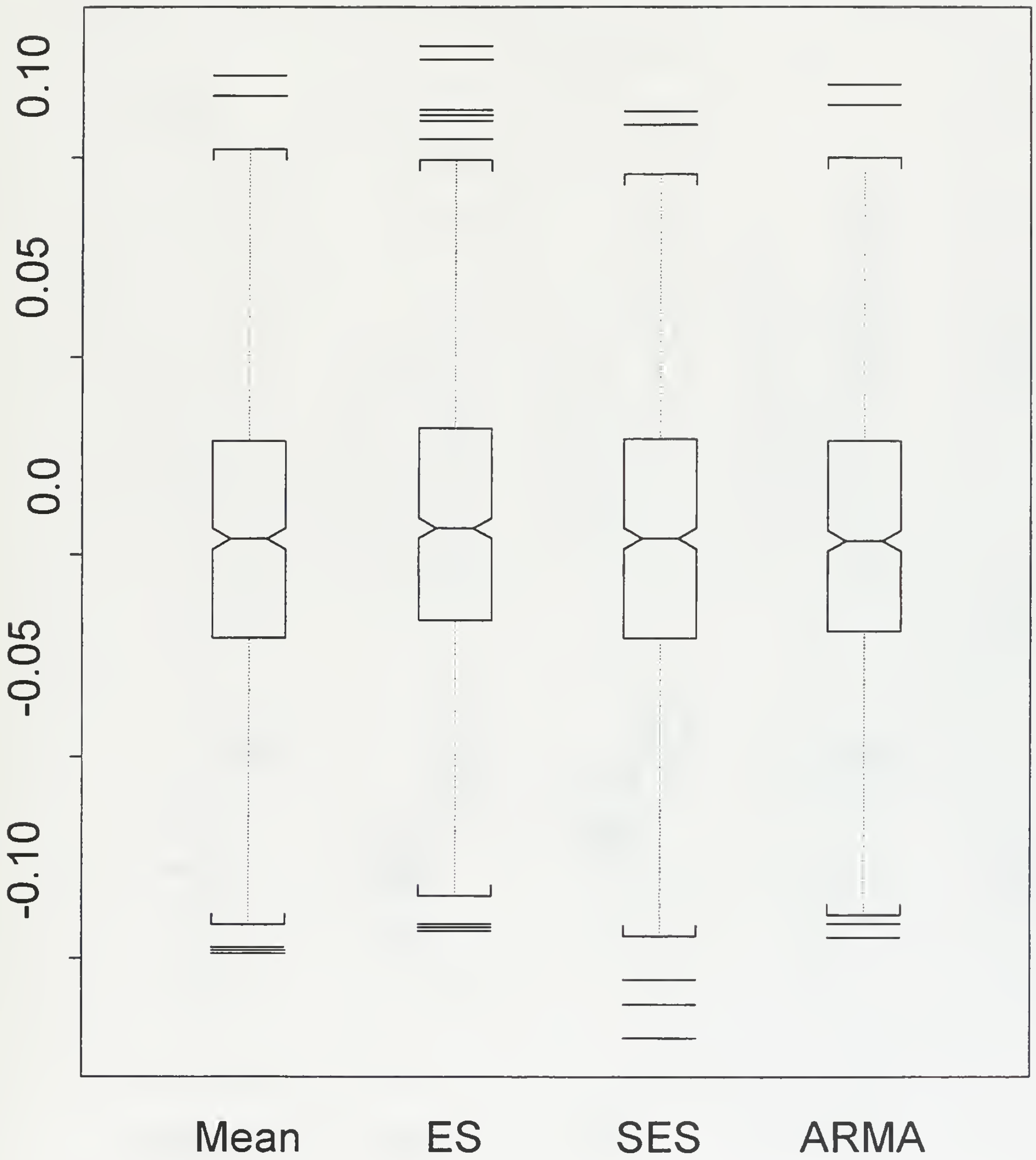

Figure G2 Relative Errors in Forecasted Strength, YOS 2.

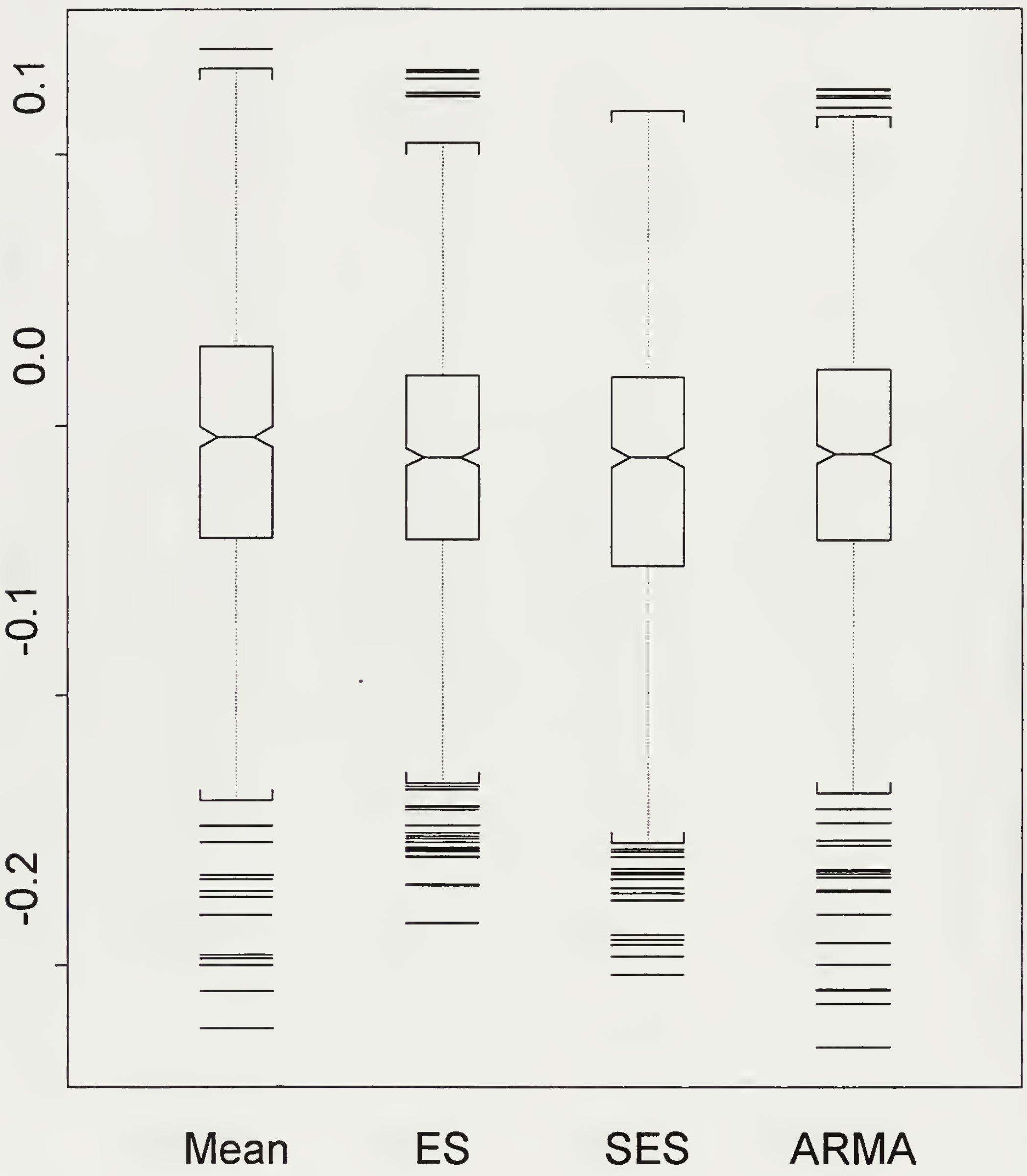

Figure G3 Relative Errors in Forecasted Strength, YOS 3.

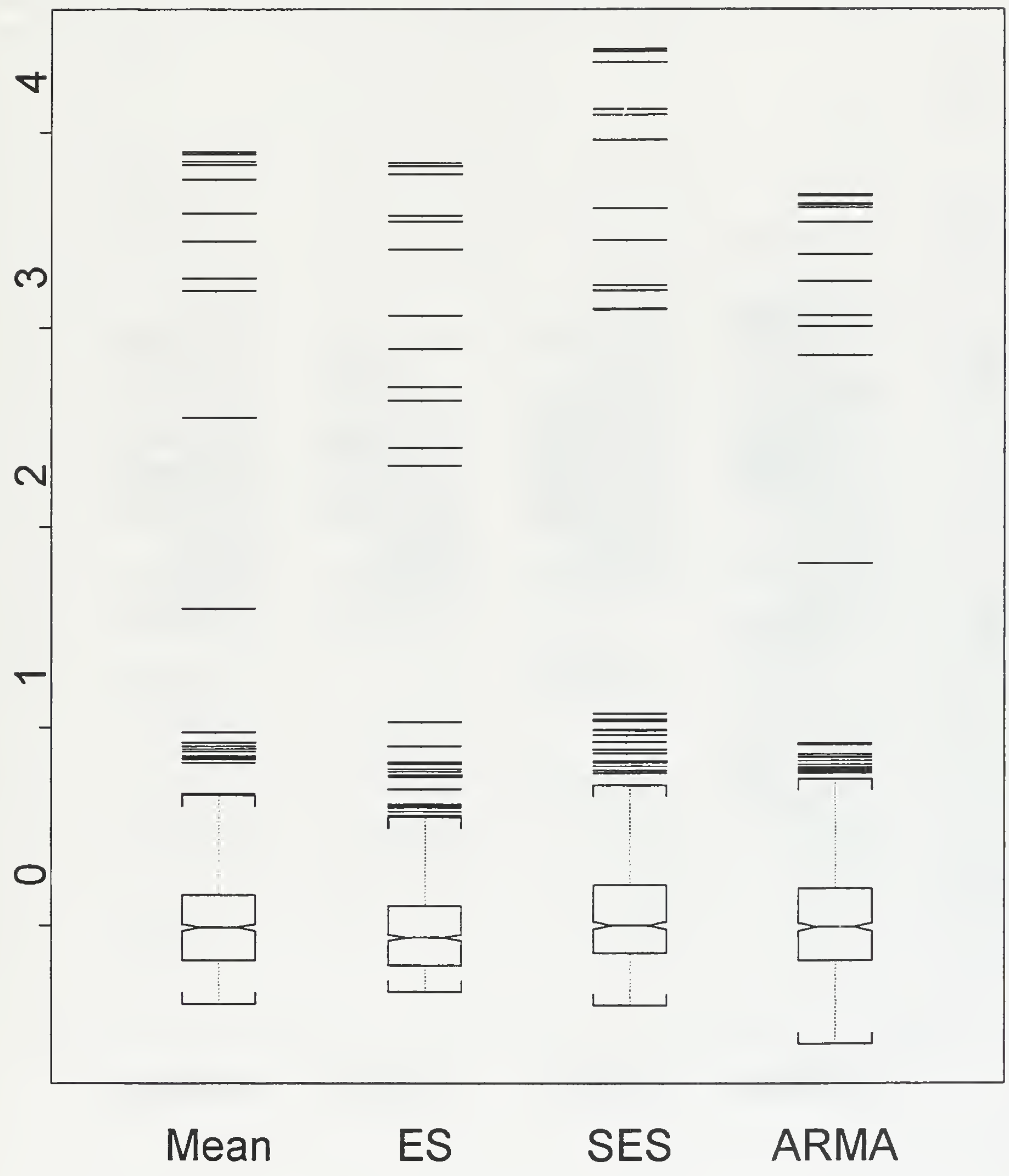

Figure G4 Relative Errors in Forecasted Strength, YOS 4.

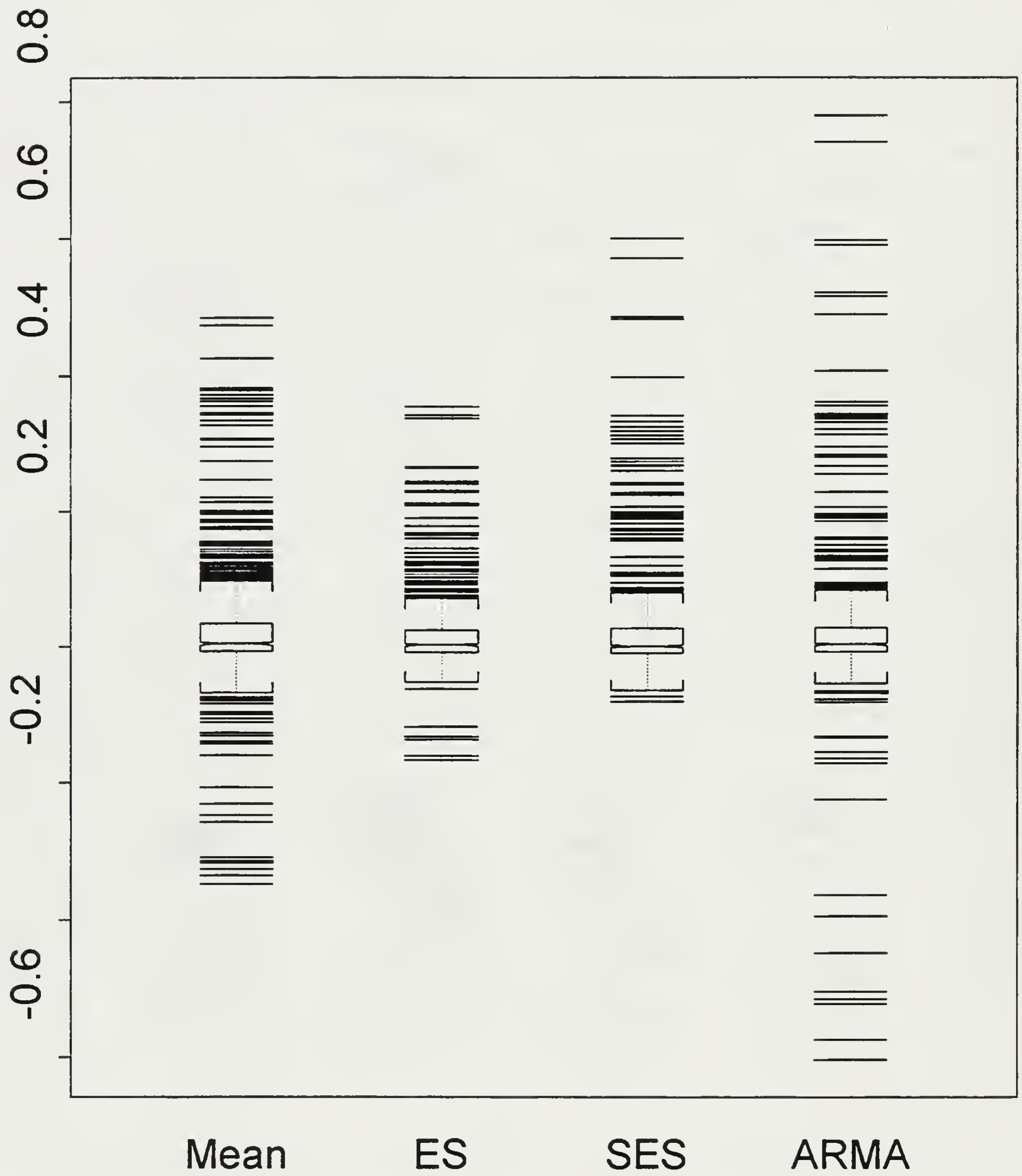

Figure G5 Relative Errors in Forecasted Strength, First Year into the Forecast Horizon, 1989.

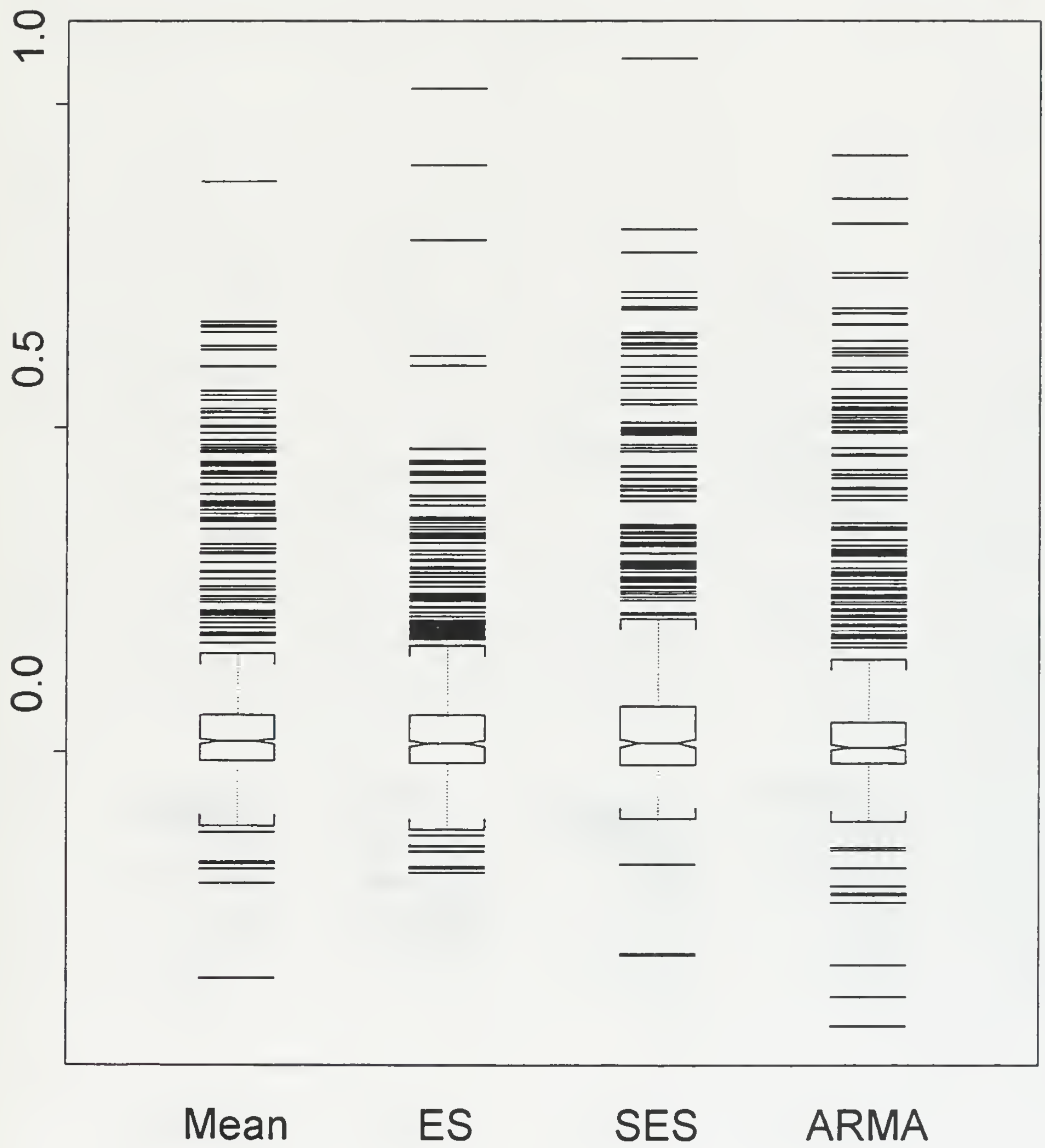

Figure G6 Relative Errors in Forecasted Strength, Second Year into the Forecast Horizon, 1990.

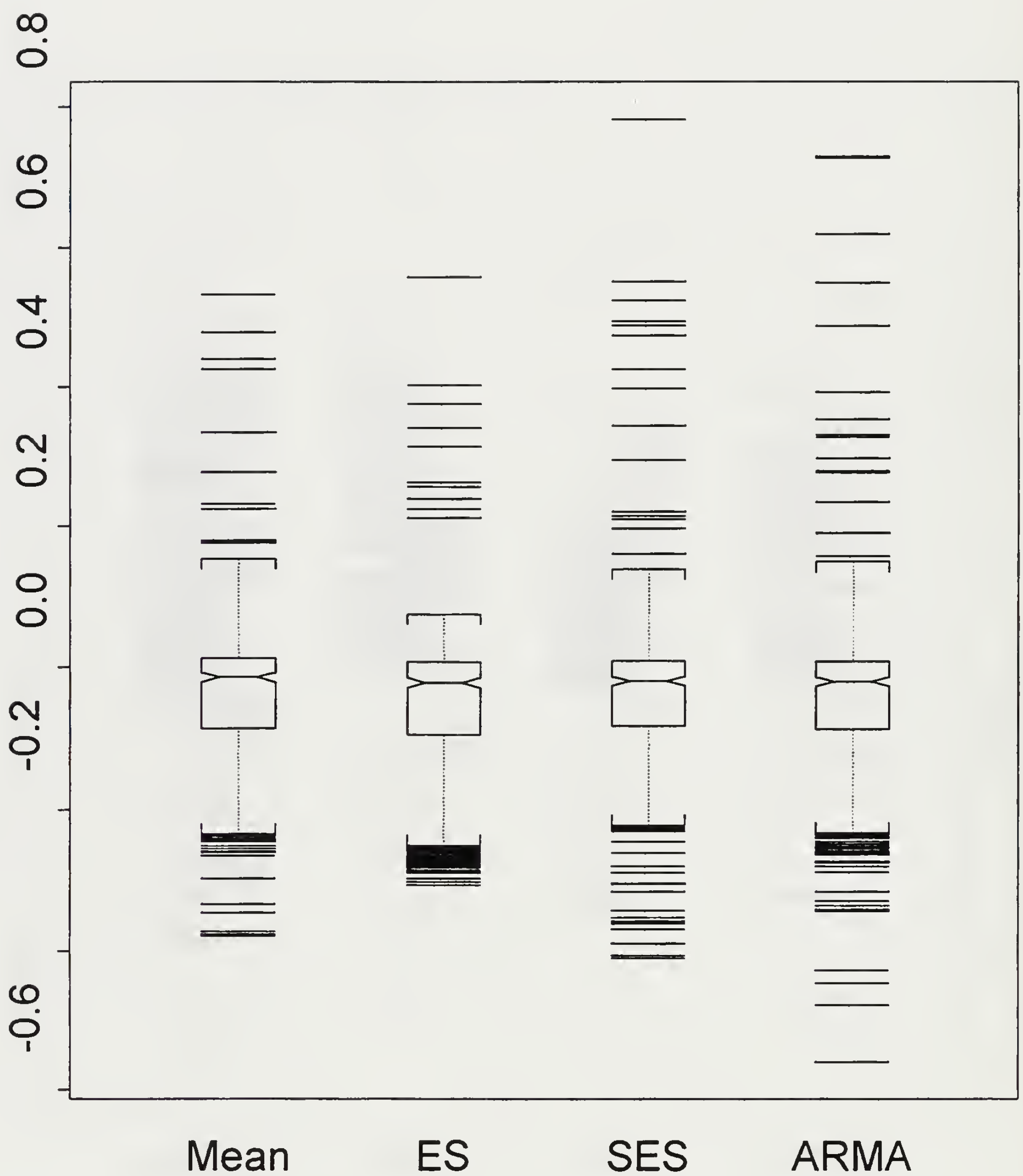

Figure G7 Relative Errors in Forecasted Strength, Third Year into the Forecast Horizon, 1991.

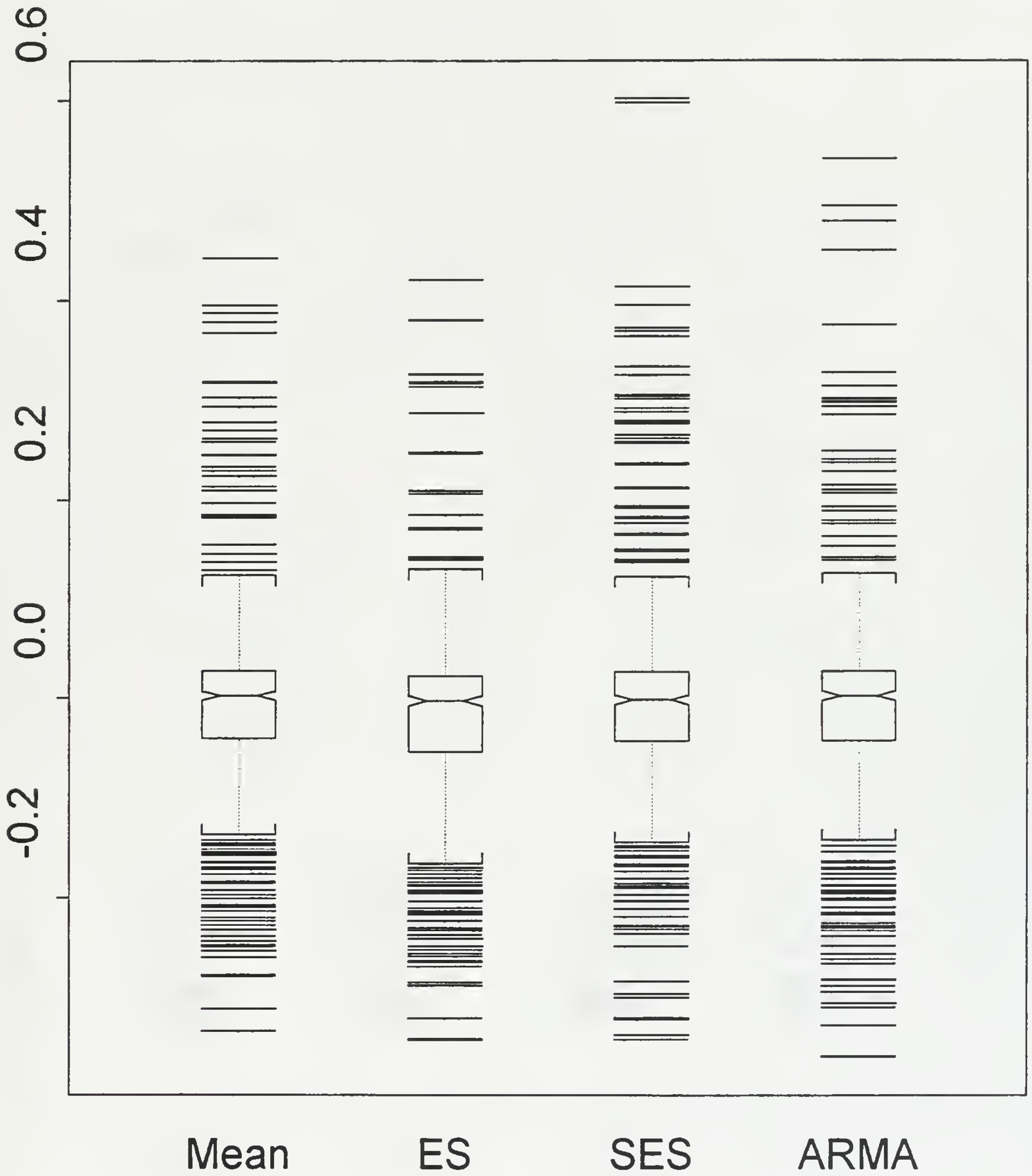

Figure G8 Relative Errors in Forecasted Strength, Fourth Year into the Forecast Horizon, 1992.

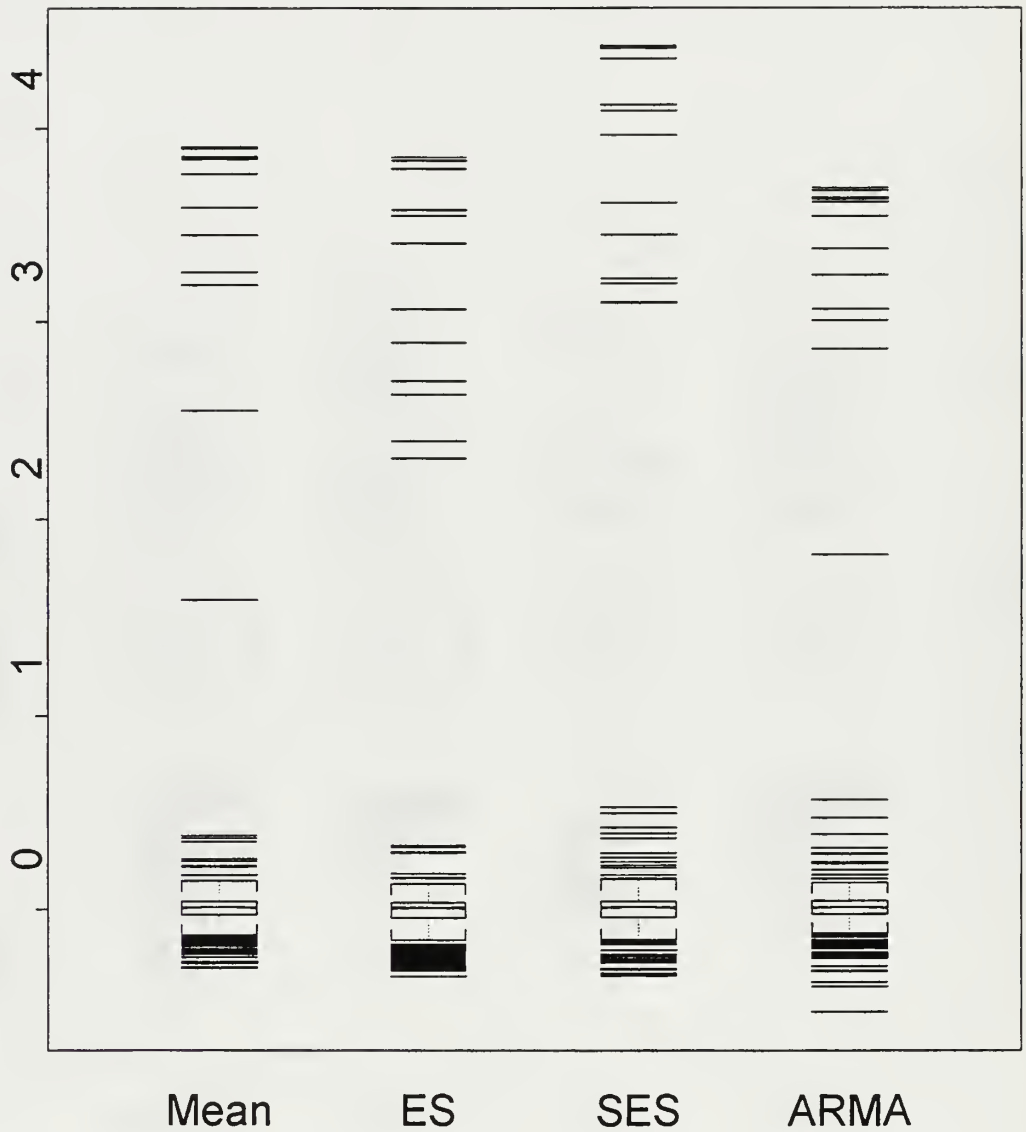

Figure G9 Relative Errors in Forecasted Strength, Fifth Year into the Forecast Horizon, 1993.

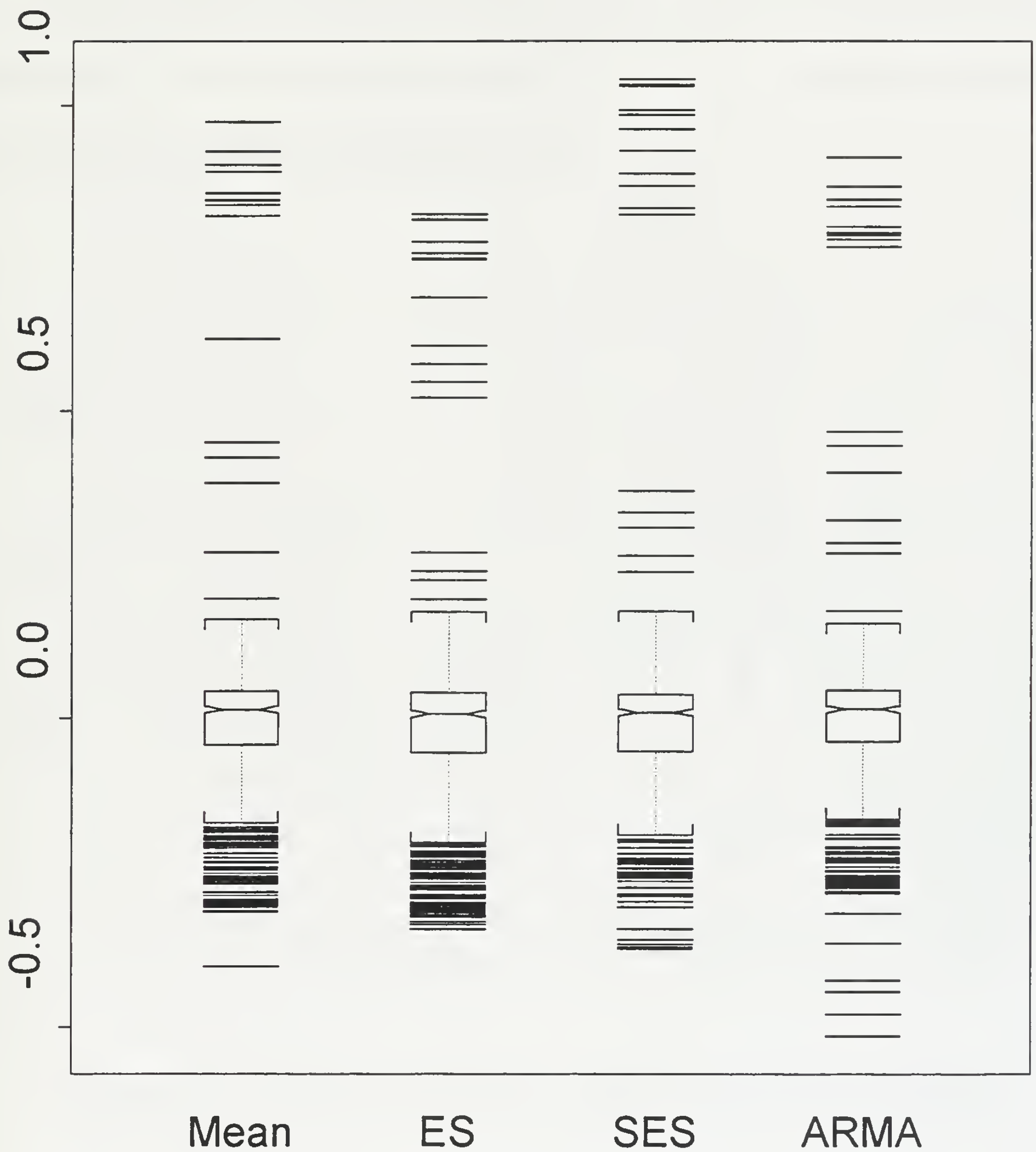

Figure G10 Relative Errors in Forecasted Strength, Sixth Year into the Forecast Horizon, 1994.

APPENDIX H. TOTAL MONTHLY ACCESSION PLOTS

This appendix contains plots of the total montly accessions into the Army from 1983 through 1994. The accession totals are grouped by month to show monthly accession trends that may in turn contribute to monthly loss trends.

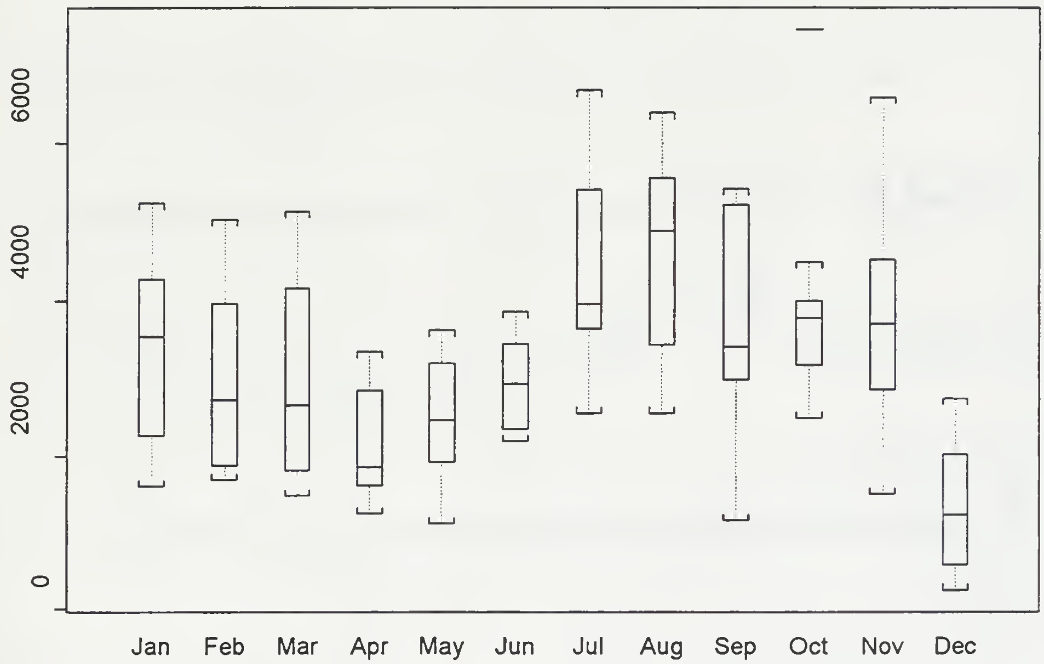

Figure H1 Accessions by month, 1983 - 1994, 3 and 4 year term enlistees.

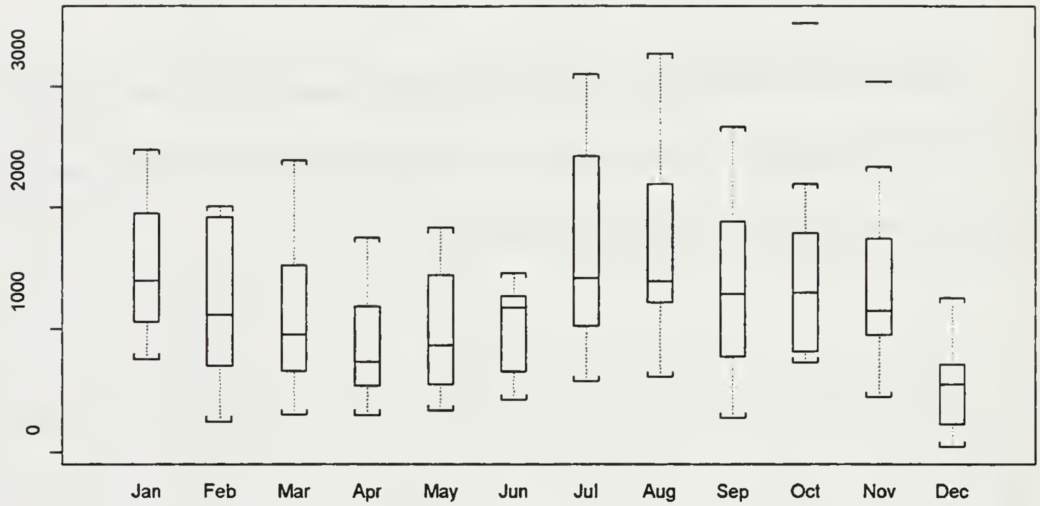

Figure H2 Accessions by month, 1983 - 1994, 3 year term enlistees only.

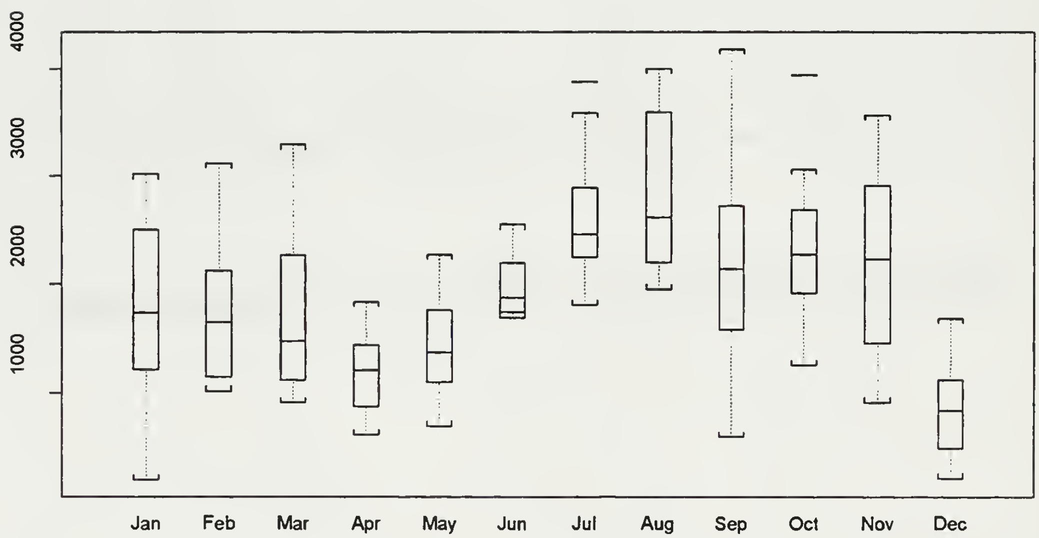

Figure H3 Accessions by month, 1983 - 1994, 4 year term enlistees only.

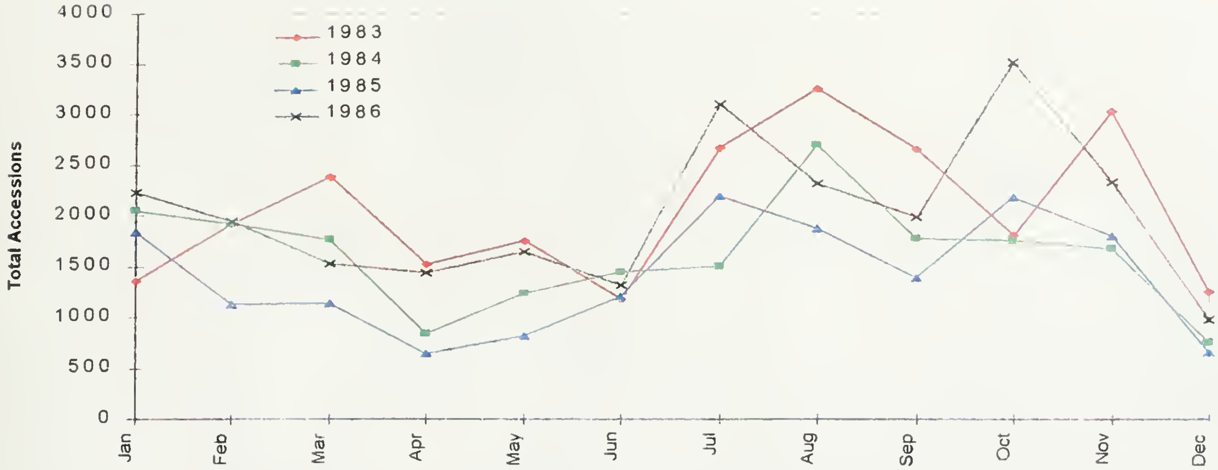

Figure H4 Total monthly accessions, 3 year enlistees, 1983 - 1986

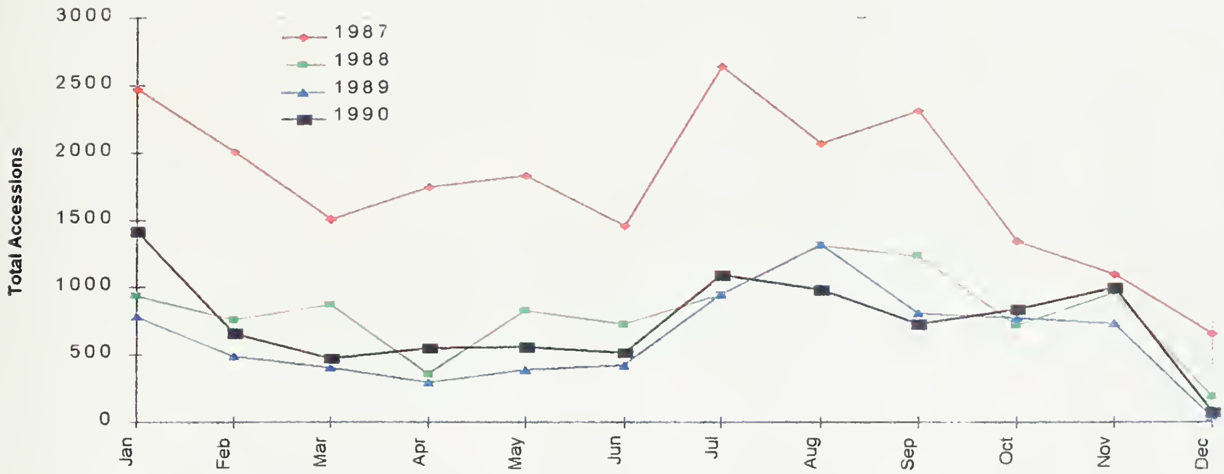

Figure H5 Total monthly accessions, 3 year enlistees, 1987 - 1990

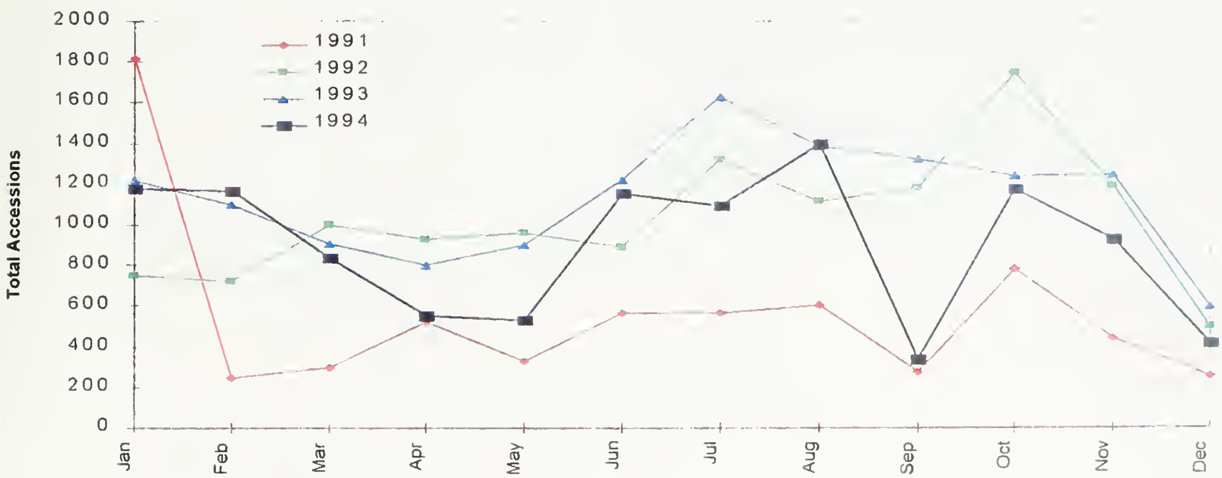

Figure H6 Total monthly accessions, 3 year enlistees. 1991 - 1994

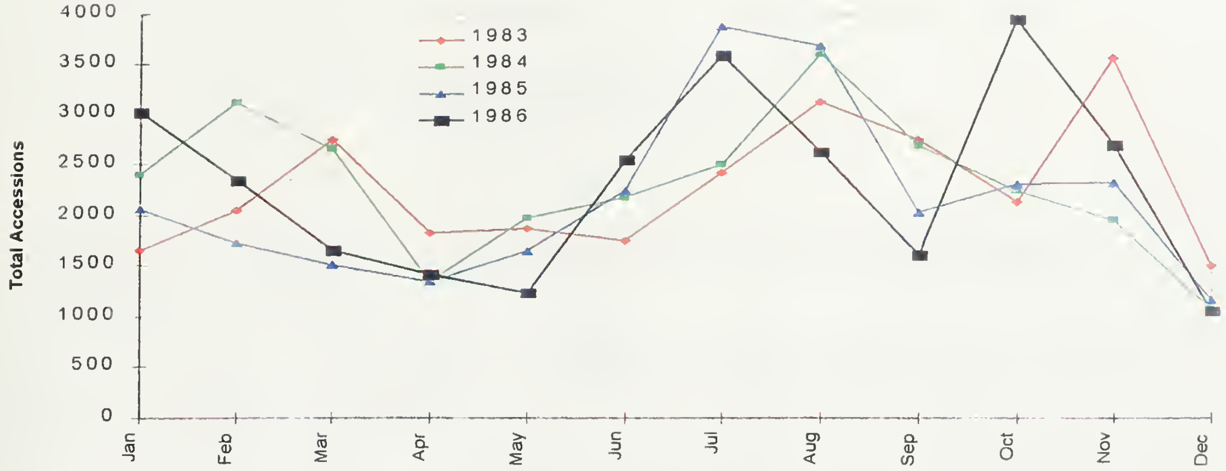

Figure H7 Total monthly accessions, 4 year enlistees, 1983 - 1986

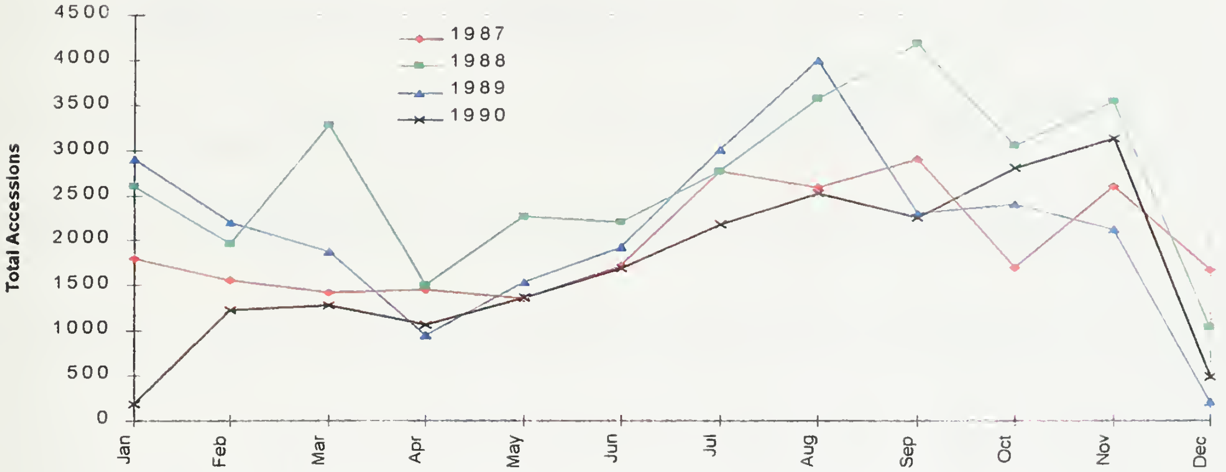

Figure H8 Total monthly accessions, 4 year enlistees, 1987 - 1990

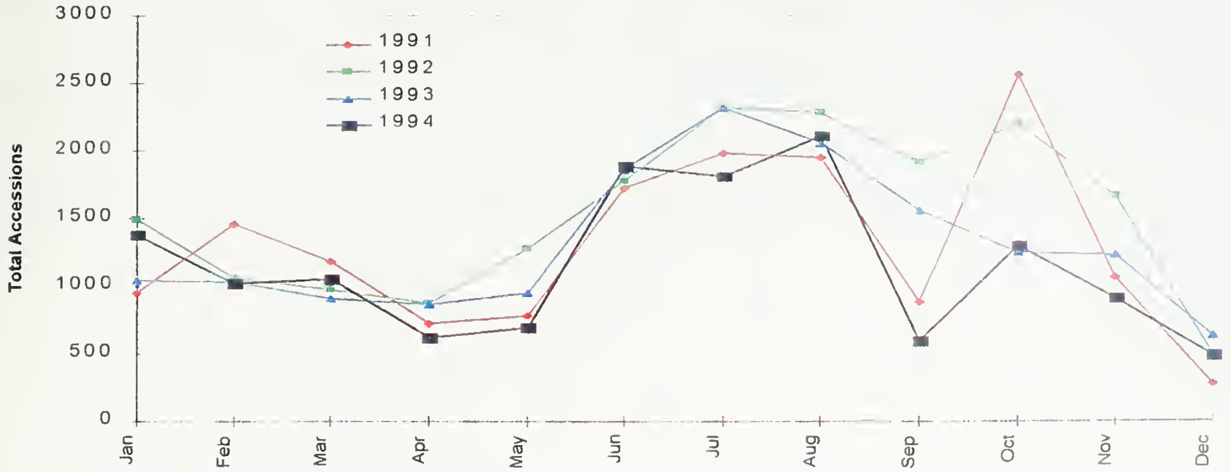

Figure H9 Total monthly accessions, 4 year enlistees, 1991 - 1994

LIST OF REFERENCES

Bartholomew, D.J. and Forbes, A.F., *Statistical Techniques for Manpower Planning*, John Wiley & Sons Inc., 1979.

Box, G.E.P. and Jenkins, G.M. , *Time Series Analysis Forecasting and Control*, Holden Day Inc., San Francisco, CA, 1976.

Dillaber, K., *ELIM-COMPLIP*, unpublished system summary maintained by ODCSPER's Female Enlisted Loss Inventory Model (FELIM) representative, 1995.

GRC, *ADP Support Services for Enlisted System: ELIM-COMPLIP System Specification*, General Research Corporation, McLean, Virginia, 1989.

GRC, *EPSMS, ELIM and MOSLS Executive Overview Briefing Slides*, General Research Corporation, McLean, Virginia, 1995.

Kalbfleisch J.D. and Prentice, R.L., *The Statistical Analysis of Failure Time Data*, John Wiley & Sons Inc., 1980.

Makridakis, S. and Wheelwright S.C., *Interactive Forcasting*, Holden-Day Inc., San Francisco, CA, 1978.

Makridakis, S. and Wheelwright S.C., *Forecasting Methods for Management*, John Wiley & Sons Inc., 1985.

Rice, J.A., *Mathematical Statistics and Data Analysis, Second Edition*, Wadsworth Inc., 1995.

S-PLUS, "Analyzing Time Series", *Guide to Statistical & Mathematical Analysis*, Math Soft Inc., 1995.

Webster's Electronic Dictionary and Thesaurus, College Edition, Random House Inc., 1992.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center 8725 John J. Kingman Rd., STE 0944 Ft. Belvoir, VA 22060-6218	2
2.	Dudlley Knox Library, Naval Postgraduate School 411 Dyer Rd. Monterey, CA 93943-5101	2
3.	Director, Training and Education MCCDC, Code C46 1019 Elliot Rd. Quantico, Virginia 22134-5027	1
4.	Director, Marine Corps Research Center MCCDC, Code C40RC 2040 Broadway Street Quantico, Virginia 22134-5107	2
5.	Director, Studies and Analysis Division MCCDC, Code C45 3300 Russell Road Quantico, Virginia 22134-5130	1
6.	Chief, Personnel Forecasting ODCSPER, Rm 2C744 300 Army Pentagon Washington, DC 20310-0300	3
7.	Professor Robert R. Read Naval Postgraduate School Code 30-360 Monterey, California 93943	2
8.	Professor Lyn R. Whitaker Naval Postgraduate School Code 30-360 Monterey, California 93943	1

9. Captain Edward T. DeWald USMC 1
5828 Jacksons Oak Court
Burke, Virginia 22015